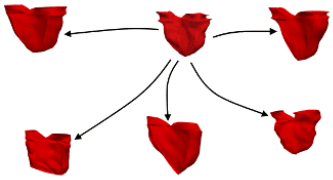
Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy - Geometrical and Statistical Methods for Modelling Biological Shape Variability

Xavier Pennec, Sarang Joshi, Mads Nielsen



MFCA 2011

3rd MICCAI workshop on

Mathematical Foundations of Computational Anatomy

Mathematical Foundations of Computational Anatomy

Geometrical and Statistical Methods for Biological Shape Variability Modeling

September 22th, 2011, Westin Harbour Castle, Toronto, Canada.

http://www.inria.fr/sophia/asclepios/events/MFCA11

Editors:
Xavier Pennec (Asclepios, INRIA Sophia-Antipolis, France)
Sarang Joshi (SCI, University of Utah, USA)
Mads Nielsen (University of Copenhagen)

informatics mathematics

MICCAI

Preface

Computational anatomy is an emerging discipline at the interface of geometry, statistics and image analysis which aims at modeling and analyzing the biological shape of tissues and organs. The goal is to estimate representative organ anatomies across diseases, populations, species or ages, to model the organ development across time (growth or aging), to establish their variability, and to correlate this variability information with other functional, genetic or structural information.

The **Mathematical Foundations of Computational Anatomy (MFCA)** workshop aims at fostering the interactions between the mathematical community around shapes and the MICCAI community in view of computational anatomy applications. It targets more particularly researchers investigating the combination of statistical and geometrical aspects in the modeling of the variability of biological shapes. The workshop is a forum for the exchange of the theoretical ideas and aims at being a source of inspiration for new methodological developments in computational anatomy. A special emphasis is put on theoretical developments, applications and results being welcomed as illustrations. Following the successful first edition of this workshop in 2006[1] and second edition in New-York in 2008[2], the third edition was held in Toronto on September 22 2011[3].

Contributions were solicited in Riemannian and group theoretical methods, geometric measurements of the anatomy, advanced statistics on deformations and shapes, metrics for computational anatomy, statistics of surfaces, modeling of growth and longitudinal shape changes. 22 submissions were reviewed by three members of the program committee. To guaranty a high level program, 11 papers only were selected for oral presentation in 4 sessions. Two of these sessions regroups classical themes of the workshop: statistics on manifolds and diffeomorphisms for surface or longitudinal registration. One session gathers papers exploring new mathematical structures beyond Riemannian geometry while the last oral session deals with the emerging theme of statistics on graphs and trees. Finally, a poster session of 5 papers addresses more application oriented works on computational anatomy.

August 2011

Xavier Pennec
General Chair
MFCA'11

[1] http://www.inria.fr/sophia/asclepios/events/MFCA06/
[2] http://www.inria.fr/sophia/asclepios/events/MFCA08/
[3] http://www.inria.fr/sophia/asclepios/events/MFCA11/

Organization

Workshop Chairs

Xavier Pennec	(INRIA Sophia-Antipolis, France)
Sarang Joshi	(SCI, University of Utah, USA)
Mads Nielsen	(University of Copenhagen, Denmark)

Program Committee

Stéphanie Allassoniere	(Ecole Polytechnique, France)
Rachid Deriche	(INRIA, France)
Stanley Durrleman	(SCI, Univ. Utah, USA)
Ian L. Dryden	(University of South Carolina, USA)
Tom Fletcher	(University of Utah, USA)
James Gee	(Univ. of Pennsylvania, USA)
Guido Gerig	(University of Utah, USA)
Polina Golland	(CSAIL, MIT, USA)
Darryl Holm	(Imperial College London, UK)
Natasha Lepore	(USC Keek School of Medicine, USA)
Yoshitaka Masutani	(University of Tokyo Hospital, Japan)
Stephen Marsland	(Massey University, New Zeland)
Michael I. Miller	(John Hopkins University, USA)
David Mumford	(Brown university, USA)
Marc Niethammer	(UNC Chapel Hill, USA)
Salvador Olmos	(University of Saragossa, Spain)
Sebastien Ourselin	(UCL, UK)
Anand Rangarajan	(University of Florida, USA)
Daniel Rueckert	(Imperial College London, UK)
Guillermo Sapiro	(University of Minnesota, USA)
Martin Styner	(UNC Chapel Hill, USA)
Anuj Srivastava	(Florida State University, USA)
Paul Thompson	(UCLA, Los-Angeles, USA)
Alain Trouvé	(ENS-Cachan, France)
Carole Twining	(University of Manchester, UK)
Baba Vemuri	(University of Florida, USA)
Francois-Xavier Vialard	(Imperial College London, UK)
William M. Wells III	(CSAIL, MIY, Boston, USA)
Laurent Younes	(John Hopkins University, USA)

Table of Contents

LDDMM and Longitudinal Deformations

Beyond Riemannian Geometry

Statistical Analysis on Riemannian Manifolds

Statistical Graphs and Trees

Poster Session

Direct LDDMM of Discrete Currents with Adaptive Finite Elements

Andreas Günther, Hans Lamecker, and Martin Weiser

Zuse Institute Berlin, Takustraße 7, D-14195 Berlin-Dahlem, Germany
guenther@zib.de

Abstract. We consider Large Deformation Diffeomorphic Metric Mapping of general m-currents. After stating an optimization algorithm in the function space of admissable morph generating velocity fields, two innovative aspects in this framework are presented and numerically investigated: First, we spatially discretize the velocity field with conforming adaptive finite elements and discuss advantages of this new approach. Second, we directly compute the temporal evolution of discrete m-current attributes.

1 Introduction

The Large Deformation Diffeomorphic Metric Mapping (LDDMM) approach initiated in [4, 18] has attracted considerable attention over the last few years in medical imaging. It allows to match highly deformed objects and as such is capable of performing inter-individual registration. LDDMM constructs a space mapping by evolving a displacement field along a velocity field, we call *wind*. Depending on the regularity of the wind, either *diffeomorphisms* [1, 12] or *homeomorphisms* [21] of the embedded space can be obtained. Thus, it provides a basis for many applications of anatomical shape analysis, where a one-to-one correspondence between different spatial objects is required.

The LDDMM technique is commonly applied for matching *currents* [5]. Currents provide a unified mathematical description of geometrical objects of dimension 0 (points), 1 (curves), 2 (surfaces) or 3 (volumes) [7, 14] which are embedded in $\mathbb{R}^3$. The spaces of m-currents are linear and equipped with an inner product and hence are a suitable tool for statistical shape analysis [5]. The induced norm provides a similarity measure for matching of source and target objects.

In [3] a particle-mesh method has been applied to 1-currents in 2D. Therein curves of same topology represented by a parameterization were approximated by a finite point set without any tangential information regardless of some geometrical level of detail for matching. Although attaching a momentum vector at each vertex of the source shape can be proven to be the optimal wind parameterization, Cotter proposes to use cubic B-splines on a fixed grid. In an enclosed efficiency discussion this sub-optimal parameterization is still competitive due to the simpler wind structure. However the numerical scheme there relies on equidistant cartesian grids with constant diffusity in order to apply FFT techniques.

In contrast to [3] we apply in 3D the *Orthogonal Matching Pursuit* (OMP) proposed in [6] to obtain a sparse representation for general m-currents at a given geometric resolution (spectral length) in terms of a sum of discrete *Dirac delta m-currents*. Currently, the LDDMM evolution of this representation is only done indirectly via an approximative scheme [5, Rem. 4.13] as depicted in Figure 1 (left) for $m = 2$, because it would otherwise require the computation of the Jacobian of the diffeomorphism, which is a challenge when discretizing the wind using Gaussian kernels [8, 9, 19]. Furthermore, one looses the connectivity of the input mesh structure in this case. However, this is not a significant problem, since the connectivity can be recovered by applying the final displacement field to the input meshes afterwards.

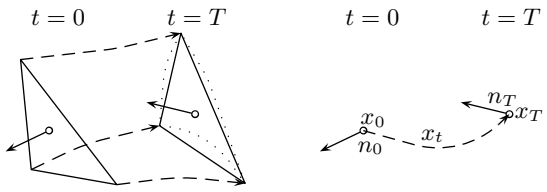


Fig. 1. Usual discrete 2-current deformation (left) versus the direct approach (right)

In this paper, we study the *direct* evolution of Dirac delta m-currents (right of Figure 1). We show that the direct approach allows to uniformly treat m-currents for $m = 0, \ldots, 3$ (Sec. 2), which to the best of our knowledge has not been shown before. We show also how to compute the Jacobian in this setting by using finite elements (FEM) to discretize the wind in the LDDMM framework (Sec. 3). Since the compactly supported basis functions are fixed in space the computation is significantly simplified. Although equidistant grids can of course be incorporated, we also in contrast to [3] consider locally refined meshes and exploit the decoupling of the wind and current discretization for an adaptive current representation, giving a significant reduction of degrees of freedom (Sec. 4). Based on our results, the increased spatial flexibility of adaptive FEM may be exploited in the future by implementing hierarchical schemes, error estimators and non-constant spatial anisotropic diffusity.

2 Continuous Matching Problem

For given shapes $\mathcal{S}, \mathcal{T} \subset \mathbb{R}^3$ we aim at constructing a sufficiently smooth bijection ϕ of $\mathbb{R}^3$ such that the distance between $\phi(\mathcal{S})$ and $\mathcal{T}$ is minimal. Here we fix the formalisms to describe the matching problem as an optimization task.

2.1 Currents

Currents are mathematical tools suited for describing geometrical objects such as points, space curves, surfaces and volumes embedded in $\mathbb{R}^3$. Their precise

definition from [7, 14] requires notation for differential forms taken from [15]. Let for $m = 0, 1, 2, 3$ the set $\mathcal{D}^m := C_c^\infty(\mathbb{R}^3, \Lambda^m\mathbb{R}^3)$ denote the vector space of all C^∞ *differential m-forms* on $\mathbb{R}^3$ with compact support. A m-current is an element of $\mathcal{D}_m$, the dual space of $\mathcal{D}^m$. The elementary *Dirac delta m-currents* $\delta_x^{u_1\wedge\ldots\wedge u_m} \in \mathcal{D}_m$ act on $\omega \in \mathcal{D}^m$ as

$$\delta_x^{u_1\wedge\ldots\wedge u_m}(\omega) = \omega(x)(u_1 \wedge \ldots \wedge u_m) \ .$$

Following the discussion in [5, Sect. 1.5.1] it turns out that for the purpose of matching currents the testspace of all C^∞ differential m-forms is not suited due to a missing bound in variation. Moreover the space $\mathcal{D}^0$ can be identified with scalar C_c^∞ functions on $\mathbb{R}^3$. For $m = 1$ and $m = 2$ the space $\mathcal{D}^m$ is isomorph to the space of vector-valued C_c^∞ functions from $\mathbb{R}^3$ to $\mathbb{R}^3$. An element of $\mathcal{D}^3$ can be written as a scalar C_c^∞ function times the determinant form on $\mathbb{R}^3$.

Both aspects motivate the use of *Reproducible Kernel Hilbert Spaces* (RKHS) W^m as testspaces.

Definition 1. *Let $d_m = 1$ for $m \in \{0, 3\}$ and $d_m = 3$ for $m \in \{1, 2\}$. For $m = 0, 1, 2, 3$ let W^m denote the dense span of d_m-vectorfields of the form $\omega(x) = k_m(x, y)a$, where $x, y \in \mathbb{R}^3$, $a \in \mathbb{R}^{d_m}$ and $k_m(x, y) = \exp(-\|x - y\|^2/\sigma_m^2)$. The space W^m can be equipped with the inner product $\langle k_m(\cdot, x)a, k_m(\cdot, y)b\rangle_{W^m} = a^* k_m(x, y)b$. Here the symbol $*$ denotes the transpose operation.*

An m-current *in $\mathbb{R}^3$ is a continuous linear functional on W^m. W_m denotes the vector space of all m-currents in $\mathbb{R}^3$.*

For $x \in \mathbb{R}^3$ and attribute *$a \in \mathbb{R}^{d_m}$ we define the elementary* Dirac delta m-currents *$\delta_x^a \in W_m$ acting on $\omega \in W^m$ as $\delta_x^a(\omega) = a^*\omega(x)$.*

The above inner product induces a norm on W^m, which can be computed efficiently via FGT even for a large number of linear combinations of the above basis functions. The chosen Gaussian kernel k_m can be considered as Green's function for some differential operator L_W (see [1, 6, 8]). With the above objects at hand the Riesz representation theorem provides a unique operator $K_W^m : W_m \to W^m$ reflecting the canonical isometry between W^m and W_m defined via

$$\langle K_W^m f, g\rangle_{W^m} = \langle f, g\rangle_{W_m, W^m} = f(g)$$

for all $f \in W_m$ and $g \in W^m$. It provides for the m-current $\mathcal{S}^m$ the Riesz representant $K_W^m\mathcal{S}^m$ as unique d_m-vectorfield on $\mathbb{R}^3$.

2.2 Homeomorphisms and Diffeomorphisms

Let Ω be an open bounded subset of $\mathbb{R}^3$ and consider functions $v_t : \bar{\Omega} \to \mathbb{R}^3$ that vanish on $\partial\Omega$. For given final time $T > 0$ and a time-dependent wind $v = (v_t)_{t\in[0,T]}$ we consider the temporal evolution of the identity map

$$\frac{\partial\phi_t^v}{\partial t} = v_t(\phi_t^v) \text{ with } \phi_0^v(x) = x \ . \tag{1}$$

In what follows it will be useful to define the *trajectory* $x_t := \phi_t^v(x)$ for some fixed space point $x \in \mathbb{R}^3$ and the map $\phi_{st}^v := \phi_t^v \circ (\phi_s^v)^{-1}$, describing the movement of a particle starting in x at time s towards $\phi_{st}^v(x)$ at time t. It is well known (see [21, Thm. C.3]), that (1) is uniquely solvable when for some $x_0 \in \Omega$ the integral $\int_0^T \|v_t(x_0)\|_{\mathbb{R}^3} + \mathrm{Lip}(v_t)\, dt$ is bounded. Furthermore its solution $\phi_t^v : \mathbb{R}^3 \to \mathbb{R}^3$ is a *homeomorphism* of Ω for all times $t \in [0,T]$. Under more restrictive assumptions onto the spatial smoothness of the wind, i.e. $v_t \in C_0^1(\Omega, \mathbb{R}^3)\ \forall t \in [0,T]$ and $\int_0^T \|v_t\|_{1,\infty}\, dt < \infty$ the unique solution of (1) is even a *diffeomorphism* of Ω for all times $t \in [0,T]$ (see [21, Thm. 8.7]). For convenience we look for the wind v_t in some Hilbert space V. Such spaces can be constructed by defining inner products associated to differential operators. Let therefore $L : V \to L^2(\mathbb{R}^3)$ be a differential operator and equip the Hilbert space V with the inner product $\langle v_t, g\rangle_V = \langle Lv_t, Lg\rangle_{L^2} = \langle L^*Lv_t, g\rangle_{V^*,V}$. Here L^* denotes the adjoint operator. For this work we use

$$S := L^*L = (-\mathrm{div}(\sigma_V^2 \nabla) + I)^k = (-\sigma_V^2 \Delta + I)^k \tag{2}$$

and $k = 1$ or $k = 2$ giving the Sobolev spaces H^k (see [8]). For given $f \in V^*$ we consider solutions $v_t \in V$ of $Sv_t = f$ with homogeneous Dirichlet boundary conditions for v_t (and v_t' if $k = 2$). Here the real parameter $\sigma_V > 0$ balances between smoothing and data fitting of the right hand side f. For other choices of L^*L and boundary conditions see [12]. Dealing with natural boundary conditions is also possible, but requires a sufficiently large domain to keep all trajectories therein. Analogous to K_W^m we introduce the isometry operator $K_V : V^* \to V$. A mathematically equivalent approach of constructing V consists in defining K_V via the Green's function $k_V(x,y)$ of L^*L, see for instance [8, 9, 19, 20].

2.3 Diffeomorphic Deformation of Currents

For $m = 0, 1, 2, 3$ let currents $\mathcal{S}^m \in W_m$ be given. Let ϕ denote a diffeomorphism on $\mathbb{R}^3$ and $d_x\phi$ the Jacobian of ϕ at x. The *pushforward* $\phi_\sharp(\mathcal{S}^m) \in W_m$ of $\mathcal{S}^m$ under ϕ is rigorously defined in [15] via the *pullback* of differential forms. For our purpose it is sufficient to mention that if $\mathcal{S}^m$ is associated to a submanifold in $\mathbb{R}^3$ its pushforward $\phi_\sharp(\mathcal{S}^m)$ under ϕ corresponds to the *deformed* submanifold $\phi(\mathcal{S}^m)$. This important property justifies to write also $\phi(\mathcal{S}^m) \in W_m$. The explicitly calculated pushforwards for elementary Dirac delta m-currents taken from [5, Table 1.2] are given in Table 1.

Table 1. Pushforwards of Dirac delta m-currents under ϕ

$m = 0$	$d_0 = 1$	$c \in \mathbb{R}$	$\phi_\sharp(\delta_x^c) = \delta_{\phi(x)}^c$
$m = 1$	$d_1 = 3$	$\tau \in \mathbb{R}^3$	$\phi_\sharp(\delta_x^\tau) = \delta_{\phi(x)}^{d_x\phi(\tau)}$
$m = 2$	$d_2 = 3$	$n \in \mathbb{R}^3$	$\phi_\sharp(\delta_x^n) = \delta_{\phi(x)}^{\lvert d_x\phi \rvert d_x\phi^{-*}(n)}$
$m = 3$	$d_3 = 1$	$\rho \in \mathbb{R}$	$\phi_\sharp(\delta_x^\rho) = \delta_{\phi(x)}^{\lvert d_x\phi \rvert \rho}$

Let some wind v be given and consider the family $(\phi_t^v)_t$ of diffeomorphisms generated via (1). The following theorem describes the direct evolution of m-current attributes $a_m \in \mathbb{R}^{d_m}$ under $(\phi_t^v)_t$, where $'$ denotes the time derivative.

Theorem 1. *The pushforwards of $\delta_{x_0}^{c_0}$, $\delta_{x_0}^{\tau_0}$, $\delta_{x_0}^{n_0}$ and $\delta_{x_0}^{\rho_0}$ under ϕ_s^v satisfying* (1) *are $\delta_{x_s}^{c_0}$, $\delta_{x_s}^{\tau_s}$, $\delta_{x_s}^{n_s}$ and $\delta_{x_s}^{\rho_s}$. Their components are given via the ODEs*

$$\begin{aligned}
x_t' &= v_t(x_t) && \text{with } x(0) = x_0\\
\tau_t' &= (d_{x_t} v_t)\tau_t && \text{with } \tau(0) = \tau_0\\
n_t' &= n_t \mathrm{tr}(d_{x_t} v_t) - (d_{x_t} v_t)^* n_t && \text{with } n(0) = n_0\\
\rho_t' &= \rho_t \mathrm{tr}(d_{x_t} v_t) && \text{with } \rho(0) = \rho_0 \ .
\end{aligned}$$

Proof. Abbreviating $J_t = d_{x_0}\phi_t^v$ and $A_t = d_{x_t} v_t$ there holds (see [1]) $J_t' = A_t J_t$ with $J(0) = I_3$. Observing the evolution of the Wronskian [13, Thm. 2.14] or via Jacobi's formula one obtains

$$|J_t|' = |J_t| \mathrm{tr}\left(J_t^{-1} J_t'\right) = |J_t| \mathrm{tr}\left(J_t^{-1} A_t J_t\right) = |J_t| \mathrm{tr}(A_t),$$

where $\mathrm{tr}(A)$ denotes the trace of a matrix A and $A^{-*} = (A^{-1})^*$. Now from Table 1 we read out

$$x_t = \phi_t^v(x_0) \ , \quad \tau_t = J_t \tau_0 \ , \quad n_t = |J_t| J_t^{-*} n_0 \ , \quad \rho_t = |J_t| \rho_0 \ .$$

Differentiation of the above equations with respect to t yields

$$\begin{aligned}
x_t' &= \phi_t^v(x_0)' = v_t(\phi_t^v(x_0)) = v_t(x_t)\\
\tau_t' &= J_t' \tau_0 = A_t J_t \tau_0 = A_t \tau_t\\
n_t' &= |J_t|' J_t^{-*} n_0 + |J_t| (J_t^{-*})' n_0 = |J_t| \mathrm{tr}(A_t) J_t^{-*} n_0 - |J_t| A_t^* J_t^{-*} n_0\\
&= n_t \mathrm{tr}(A_t) - A_t^* n_t\\
\rho_t' &= |J_t|' \rho_0 = |J_t| \mathrm{tr}(A_t) \rho_0 = \rho_t \mathrm{tr}(A_t),
\end{aligned}$$

which proves the assertion. □

Remark 1. The authors emphasize the striking advantage that Theorem 1 enables to find the final position and attribute of a Dirac delta m-current without computing the Jacobian of the deformation. The appearing ODEs only involve the Jacobian of the velocity fields, which will be given in a closed form in any case.

2.4 Optimization Problem in Function Space

Let source $\mathcal{S}^m \in W_m$ and target current $\mathcal{T}^m \in W_m$ be given for $m = 0, \dots, 3$. For given wind v we define the deformed current $\mathcal{S}_t^m := \phi_t^v(\mathcal{S}^m)$ at time t. *Matching* means the minimization of the distance of the deformed source current at final time $\mathcal{S}_T^m$ with its target current $\mathcal{T}^m$, i.e. minimizing the dual norm $\|\phi_T^v(\mathcal{S}^m) - \mathcal{T}^m\|_{W_m} = \|\mathcal{S}_T^m - \mathcal{T}^m\|_{W_m}$ in the space of m-currents.

Given a regularization parameter $\gamma > 0$ and matching weights $\omega_m \geq 0$ we consider for $v \in L^2([0,T],V)$ the following optimization problem:

$$J(v) := \gamma \int_0^T \|v_t\|_V^2 \, dt + \sum_{m=0}^{3} \omega_m \|\phi_T^v(\mathcal{S}^m) - \mathcal{T}^m\|_{W_m}^2 \to \min \quad . \tag{3}$$

Here the first summand involves the kinetic energy of the wind. The existence of a solution for (3) is proven in [10], however it is generally not unique [2]. Following [9] the gradient of J in $L^2([0,T],V)$ at fixed v is given by $(\nabla J)_t = 2\gamma v_t + 2K_V(f_t)$, where $f_t \in V^*$ is defined by

$$f_t(u) = \sum_{m=0}^{3} \omega_m \langle \mathcal{S}_t^m, \nabla (K_W^m (\mathcal{S}_T^m - \mathcal{T}^m) \circ \phi_{tT}^v)^* u \rangle_{W_m, W^m} \quad \forall u \in V \quad .$$

For further discussion concerning the choice of the gradients metric we refer the reader to [1]. With the above quantities at hand one is able to state a steepest descent optimization algorithm in the function space of velocity fields v.

3 Discrete Matching Problem

3.1 Discretization of the Wind by Finite Elements (FE)

In the field of optimal current matching mainly wind discretizations of the form

$$v_t(x) = \sum\nolimits_j k_V(x_{j,t}, x)\alpha_{j,t} \tag{4}$$

have been considered. Here $\alpha_{j,t} \in \mathbb{R}^3$ are the time-dependent momentum vectors and k_V denotes a Gaussian kernel with some global kernel parameter $\sigma_V > 0$, describing the coherent movement of neighboring particles. In order to apply *Fast Gauss Transform* (FGT) for efficient evaluation, σ_V is necessarily a constant. Although (4) can be proven to be the optimal wind parameterization, the spatial movement of non-compactly supported basis functions along trajectories $x_{j,t}$ may cause numerical difficulties. Too small distances between them cause a redundant or badly conditioned description of the velocity field while the absence of trajectories in a part of the domain produces almost no wind there for small kernel sizes. The trajectory density varies during optimization and hence is difficult to control. Because the trajectories' starting points are the spatial components of the Dirac delta source currents the number of trajectories is fixed and hence a notion of adaptivity for the velocity field can hardly be introduced. Finally, as mentioned in Sect. 2.2, C^∞ smoothness is not required to solve the evolution equation.

In [16] and [17], some of the above mentioned drawbacks are overcome by incorporating multiple kernel shapes at different scales σ_V.

Similar to the particle-mesh method proposed in [3], we follow another suboptimal approach completely decoupling the discretization of the space of m-currents W_m from the spatial velocity space V. Keeping in mind that fast point

evaluation of the wind is essential for performance, we consider adaptive hexahedral grids for Ω with hanging nodes saved as an octree. Over such hexahedral grids we construct either C^1 conforming Hermite finite elements of third order or simpler C^0 conforming Lagrange finite elements of first order. The wind for fixed time $t \in [0,T]$ in the FE basis $\{\varphi_j\}_j$ takes the form

$$v_t(x) = \sum\nolimits_{j=1}^{n} \varphi_j(x)\alpha_{j,t} \ . \tag{5}$$

In contrast to radial basis functions, locally constant functions are contained in the ansatz space and allow to represent local or even global translations with few *degrees of freedom* (DOF). Due to the compactly supported basis functions there is no need for an approximate evaluation like FGT with further unknown tolerance parameters. Since the basis functions are fixed in space, the underlying mesh provides a natural clustering which can be exploited via a smart parallel octree search algorithm for point evaluation. Furthermore this approach provides a multilevel wind hierarchy with a fraction of DOFs on the coarsest mesh level completely decoupled from the m-current discretization. These advantages also appear in the particle-mesh method with tensor-products of cubic B-splines for instance. But since we do not apply FFT for wind evaluation, we are more flexible with adaptive meshes and do not require a box domain. Moreover non-constant anisotropic diffusity $\sigma_V(x) \in \mathbb{R}^{3\times 3}$ may be incorporated in future.

A difficulty arises in the computation of the $L^2([0,T],V)$-gradient. It permanently involves the solution of a second ($k=1$) or fourth ($k=2$) order elliptic PDE in every time-step and every iteration. It is clear that one should employ suited preconditioners and / or multigrid solvers. Using existing FE libraries limits its implementation overhead. We chose libMesh [11], which provides conforming C^1 finite elements on adaptive hexahedral meshes.

The development of adaptive mesh refinement is beyond the scope of this work. Here, we provide a proof of concept that adaptive grids can easily be incorporated. Therefore we simply geometrically refine near $\mathcal{S}^m \cup \mathcal{T}^m$ considered as subsets of $\mathbb{R}^3$ equally for all times. More sophisticated error indicators suggesting refinements could be the scalar fields $|K_W^m(\mathcal{S}_T^m - \mathcal{T}^m) \circ \phi_{tT}^v|$, $|v_t|$ or $|L^*Lv_t|$. The latter one measures the smoothness of v_t. Moreover thinking of hierarchical error estimators one could compute $\|v_t^1 - I_h v_t^1\|_{L^2(Q)}$ or even $\|v_t^1 - v_t^0\|_{L^2(Q)}$ on hexahedrons Q, where v^i denote the numerical solutions for the C^i conforming FE discretization and I_h is the usual Lagrange interpolation operator.

All appearing ODEs are numerically integrated via the explicit method of Heun on an equidistant decomposition of the time interval $[0,T]$.

3.2 Current Compression and Direct Evolution

For approximating a m-current $\mathcal{S}^m \in W_m$ as $\hat{\mathcal{S}}^m = \sum_{i=1}^{s_m} \delta_{x_i}^{a_i} \in W_m$ we use the *Orthogonal Matching Pursuit* (OMP) proposed in [6]. This method iteratively selects the most important points x_i and computes corresponding attributes a_i (i.e. c_i, τ_i, n_i, ρ_i) of a general m-current via a greedy algorithm. It has the advantage of compressing the current information for a characteristical spectral

length $\sigma_m > 0$ towards a fraction. This enables the design of highly efficient numerical solution algorithms. The approximation error in OMP is controlled by a threshold parameter and the grid size of a uniform testgrid.

The obvious drawback of loosing the connectivity between vertices (for $m \geq 1$) can be compensated by applying the obtained optimal diffeomorphism to all connected vertices whenever it is required. This only requires one additional forward flow computation at the end.

In [19, Sec. 3.2] two methods to deform a 2-current $\hat{\mathcal{S}}^2$ under a family of diffeomorphisms $(\phi_t^v)_t$ are described. In contrast to all previous work, we will pursue the *direct* approach motivated by Theorem 1. For 2-currents, only 1 instead of 3 trajectories is needed to evolve the normal n_0 (Fig. 1). In general, the direct approach requires only one trajectory per attribute, hence decreasing the number of variables in the computation, whereas in the indirect case an artificial m-simplex with $m+1$ vertices is attached.

Remark 2. To quote Rem. 4.13 in [5] the direct evolution of current attributes is closer to the analytical concept of currents and is particularly suited for OMP, where no connectivity between the points is provided. But [5] indicates the need of Jacobi matrices (as they arise in Theorem 1) as a disadvantage for numerical implementation. At least for the gradient computation in the next section we benefit from the simpler structure of v_t in (5), which in Lemma 2 enables easy evaluations of $d_{x_t} v_t = \sum_{j=1}^n \alpha_{j,t} \nabla \varphi_j(x_t)^*$ and hence $\operatorname{tr}(d_{x_t} v_t)$, $(d_{x_t} v_t)w$ and $(d_{x_t} v_t)^* w$ for a vector $w \in \mathbb{R}^3$. Note that all sums over j are *local* sums due to the compact support of the basis functions φ_j.

3.3 Discrete Optimization Problem and its Gradient

Let $a(\cdot,\cdot)$ denote the bilinear form corresponding to the elliptic differential operator S from (2). We define the sparse symmetric, positive definite matrix $\mathbf{S} := [a(\varphi_i, \varphi_j)]_{i;j=1}^n$ using the FE basis $\{\varphi_j\}_j$ from Sect. 3.1. Moreover we introduce the block vectors $\boldsymbol{\alpha}_t := [\alpha_{i,t}]_{i=1}^n$, $\mathbf{x}_t := \mathbf{x}_t^m := [x_{i,t}]_{i=1}^{s_m}$ and $\mathbf{a}_t := \mathbf{a}_t^m := [a_{i,t}]_{i=1}^{s_m}$. This notation allows to write the matching terms as

$$E^m = E^m(\mathbf{x}_T, \mathbf{a}_T) = \|\phi_T^v(\hat{\mathcal{S}}^m) - \hat{\mathcal{T}}^m\|_{W_m}^2 = \|\textstyle\sum_{i=1}^{s_m} \delta_{x_{i,T}}^{a_{i,T}} - \sum_{j=1}^{r_m} \delta_{y_j}^{b_j}\|_{W_m}^2 \ .$$

Finally the discrete form of the current matching problem (3) is

$$\hat{J}(\boldsymbol{\alpha}_t) := \gamma \int_0^T \|v_t(\boldsymbol{\alpha}_t)\|_V^2 \, dt + \sum_{m=0}^3 \omega_m \|\phi_T^v(\hat{\mathcal{S}}^m) - \hat{\mathcal{T}}^m\|_{W_m}^2 \to \min$$

or even shorter via (5) and $\|v_t(\boldsymbol{\alpha}_t)\|_V^2 = a(v_t(\boldsymbol{\alpha}_t), v_t(\boldsymbol{\alpha}_t)) = \boldsymbol{\alpha}_t^* \mathbf{S} \boldsymbol{\alpha}_t$

$$\hat{J}(\boldsymbol{\alpha}_t) = \gamma \int_0^T \boldsymbol{\alpha}_t^* \mathbf{S} \boldsymbol{\alpha}_t \, dt + \sum_{m=0}^3 \omega_m E^m(\mathbf{x}_T, \mathbf{a}_T) \to \min \ . \tag{6}$$

The analytical computation of the gradient at given $\boldsymbol{\alpha}_t$ becomes manageable though the simpler wind representation (5). Numerically the computation is

more involved due to presence of Hessians of basis functions. But these are easily provided via the already mentioned libMesh library.

Theorem 2. *The gradient of $\hat{J}$ in the L^2-metric is*

$$(\nabla \hat{J})_t = 2\gamma \mathbf{S}\boldsymbol{\alpha}_t + \sum_{m=0}^{3} \omega_m((\boldsymbol{\varphi}_t^m)^* \boldsymbol{\eta}_t^m + (\partial_{\boldsymbol{\alpha}} \mathbf{g}_t^m)^* \boldsymbol{\zeta}_t^m), \tag{7}$$

$$\begin{aligned} \textit{with} \qquad \boldsymbol{\varphi}_t^m &= [\varphi_j(x_{i,t}) I_{d_m}]_{i=1\dots s_m; j=1\dots n} \\ \boldsymbol{\zeta}_t^m &= \nabla_{\mathbf{a}_T} E^m + \int_t^T (\partial_{\mathbf{a}} \mathbf{g}_s^m)^* \boldsymbol{\zeta}_s^m \, ds \\ \boldsymbol{\eta}_t^m &= \nabla_{\mathbf{x}_T} E^m + \int_t^T (\partial_{\mathbf{x}} \mathbf{g}_s^m)^* \boldsymbol{\zeta}_s^m \, ds \ . \end{aligned}$$

The proof is postponed to the appendix. The remaining quantities $\nabla_{\mathbf{x}_T} E^m, \nabla_{\mathbf{a}_T} E^m, \partial_{\boldsymbol{\alpha}} \mathbf{g}_t^m, \partial_{\mathbf{x}} \mathbf{g}_t^m$ and $\partial_{\mathbf{a}} \mathbf{g}_t^m$ from Theorem 2 for each m are specified in the next two lemmas.

Lemma 1. *Let $f_m(x) = \sum_{i=1}^{s_m} k_m(x_{i,T}, x) a_{i,T} - \sum_{j=1}^{r_m} k_m(y_j, x) b_j$. There hold*

$$\nabla_{\mathbf{x}_T} E^m = [2(d_{x_{i,T}} f_m(x_{i,T}))^* a_{i,T}]_{i=1}^{s_m} \quad \textit{and} \quad \nabla_{\mathbf{a}_T} E^m = [2 f_m(x_{i,T})]_{i=1}^{s_m} \ .$$

Proof.

$$\begin{aligned} (\partial_{x_T} E^m)\eta &= 2\left[\left(\partial_{x_T} \textstyle\sum_{i=1}^{s_m} \delta_{x_{i,T}}^{a_{i,T}}\right)\eta\right](f_m) = 2a_T^* \left(d_{x_T} f_m(x_T)\right) \eta \\ \nabla_{x_T} E^m &= 2(d_{x_T} f_m(x_T))^* a_T \\ &= 2\Big(\textstyle\sum_{i=1}^{s_m} (\nabla_2 k_m(x_{i,T}, x_T)) a_{i,T}^* - \sum_{j=1}^{r_m} (\nabla_2 k_m(y_j, x_T)) b_j^* \Big) a_T \\ (\partial_{a_T} E^m)\eta &= 2\left[\left(\partial_{a_T} \textstyle\sum_{i=1}^{s_m} \delta_{x_{i,T}}^{a_{i,T}}\right)\eta\right](f_m) = 2\eta^* f_m(x_T) \ . \end{aligned}$$

□

Lemma 2. *For $\mathbf{g}_t^m$ in (11) their sparse Jacobians are given via*

$$\begin{aligned} \mathbf{g}_t^0 &= \mathbf{0} \\ \partial_{\boldsymbol{\alpha}} \mathbf{g}_t^1 &= \left[(\tau_{i,t}^* \nabla \varphi_j(x_{i,t})) I_3\right]_{i=1\dots s_1; j=1\dots n} \\ \partial_{\boldsymbol{\alpha}} \mathbf{g}_t^2 &= \left[n_{i,t} \nabla \varphi_j(x_{i,t})^* - \nabla \varphi_j(x_{i,t}) n_{i,t}^*\right]_{i=1\dots s_2; j=1\dots n} \\ \partial_{\boldsymbol{\alpha}} \mathbf{g}_t^3 &= [\rho_{i,t} \nabla \varphi_j(x_{i,t})^*]_{i=1\dots s_3; j=1\dots n} \\ \partial_{\mathbf{x}} \mathbf{g}_t^1 &= \operatorname{diag}\Big[\textstyle\sum_{j=1}^{n} \alpha_{j,t} \tau_{i,t}^* H_{\varphi_j}(x_{i,t}) \Big]_{i=1}^{s_1} \\ \partial_{\mathbf{x}} \mathbf{g}_t^2 &= \operatorname{diag}\Big[\textstyle\sum_{j=1}^{n} n_{i,t} (\alpha_{j,t}^* H_{\varphi_j}(x_{i,t})) - (\alpha_{j,t}^* n_{i,t}) H_{\varphi_j}(x_{i,t}) \Big]_{i=1}^{s_2} \\ \partial_{\mathbf{x}} \mathbf{g}_t^3 &= \operatorname{diag}\Big[\textstyle\sum_{j=1}^{n} \rho_{i,t} \alpha_{j,t}^* H_{\varphi_j}(x_{i,t}) \Big]_{i=1}^{s_3} \\ \partial_{\boldsymbol{\tau}} \mathbf{g}_t^1 &= \operatorname{diag}\Big[\textstyle\sum_{j=1}^{n} \alpha_{j,t} \nabla \varphi_j(x_{i,t})^* \Big]_{i=1}^{s_1} \end{aligned}$$

$$\partial_{\mathbf{n}} \mathbf{g}_t^2 = \operatorname{diag}\Big[\textstyle\sum_{j=1}^n (\alpha_{j,t}^* \nabla\varphi_j(x_{i,t})) I_3 - \nabla\varphi_j(x_{i,t}) \alpha_{j,t}^* \Big]_{i=1}^{s_2}$$
$$\partial_{\boldsymbol{\rho}} \mathbf{g}_t^3 = \operatorname{diag}\Big[\textstyle\sum_{j=1}^n \alpha_{j,t}^* \nabla\varphi_j(x_{i,t}) \Big]_{i=1}^{s_3},$$

where $H_{\varphi_j}(x_{i,t})$ *denote the Hessian of* φ_j *at* $x_{i,t}$.

Proof. The proof for all cases of m can easily be adapted from the case $m = 2$. For this choice the derivatives of $\mathbf{g}_t^2$ follow from direct calculations starting with

$$\begin{aligned}
\mathbf{g}_t^2 &= \mathbf{g}^2(\boldsymbol{\alpha}_t, \mathbf{x}_t, \mathbf{n}_t) = \operatorname{diag}\left[\operatorname{tr}(d_{x_{i,t}} v_t) I_3 - (d_{x_{i,t}} v_t)^*\right]_{i=1}^{s_2} \mathbf{n}_t \\
&= \operatorname{diag}\Big[\textstyle\sum_{j=1}^n (\alpha_{j,t}^* \nabla\varphi_j(x_{i,t})) I_3 - \nabla\varphi_j(x_{i,t}) \alpha_{j,t}^* \Big]_{i=1}^{s_2} \mathbf{n}_t \\
&= \Big[\textstyle\sum_{j=1}^n n_{i,t} (\alpha_{j,t}^* \nabla\varphi_j(x_{i,t})) - \nabla\varphi_j(x_{i,t}) (\alpha_{j,t}^* n_{i,t}) \Big]_{i=1}^{s_2},
\end{aligned}$$

where $\operatorname{diag}[\mathbf{v}] = [\delta_{ij} v_i]_{i,j=1}^s$ for $\mathbf{v} \in \mathbb{R}^s$ and δ_{ij} denotes the Kronecker delta. □

Corollary 1. *If* $\omega_m = 0$ *for* $m > 0$ *Theorem 2 simply provides*

$$(\nabla \hat{J})_t = 2\gamma \mathbf{S} \boldsymbol{\alpha}_t + \omega_0 (\boldsymbol{\varphi}_t^0)^* \nabla_{\mathbf{x}_T} E^0 \quad .$$

Remark 3. The $L^2([0,T], V)$-gradient of $\hat{J}$ is immediately obtained by applying $\mathbf{S}^{-1}$ from the left in equation (7).

4 Numerical Experiments

Since the numerical implementation is not yet fully tested, we postpone the investigation of the cases $m > 0$ and only consider the case $m = 0$, i.e. $\omega_i = \delta_{i0}$. The surfaces $\mathcal{S}$ and $\mathcal{T}$ are depicted in Fig. 2. To both of them we apply the OMP with $\sigma_0 = 8$ towards $\hat{\mathcal{S}}$ with $s_0 = 1746$ points and $\hat{\mathcal{T}}$ with $r_0 = 2141$ points, which are sketched as set of spheres of diameter 8 in Fig. 3.

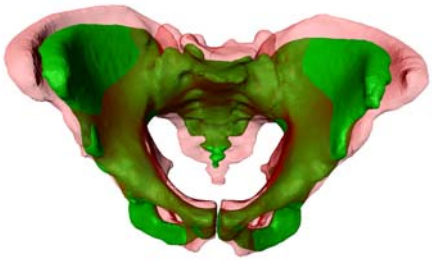


Fig. 2. $\mathcal{S}$ (dark green), $\mathcal{T}$ (light red)

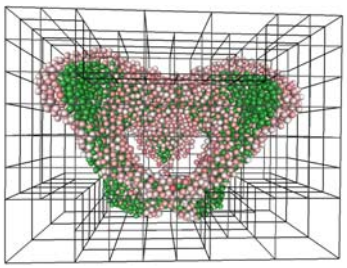


Fig. 3. $\hat{\mathcal{S}}$ (green), $\hat{\mathcal{T}}$ (red), grid

We solve the discrete matching problem (6) on $\Omega = (0, 346.56) \times (0, 205.76) \times (0, 256.96)$ with $\gamma = 0$ and tuned σ_V. For Lagrange FE ($k = 1$) on a hexahedral

adaptive grid from Fig. 3 having 568 nodes (133 of them are hanging nodes) we choose $\sigma_V = 100$. The result $\mathcal{S}_T$ is shown in Fig. 4. Secondly we solve problem (6) with Hermite FE ($k = 2$) on a uniform coarse grid having 120 nodes, whose solution is shown in Fig. 5 for $\sigma_V = 15$. Finally in Fig. 6 we compare our results with the software ExoShape[1], generating C^∞ wind via ansatz (4) with $\sigma_V = 30$. The greyscale highlights the term $\mathrm{dist}_{x\in\mathcal{S}_T}(x, \mathcal{T})$. One should keep in mind, that with $\gamma = 0$ the deformation norm vanishes in the example, but different differential operators are still present for velocity field evaluation. Although a proper comparison between all methods should use the same deformation norm $\|\cdot\|_V$, Exoshape does not easily support its change.

Fig. 4. $\mathcal{S}_T$ for C^0 wind

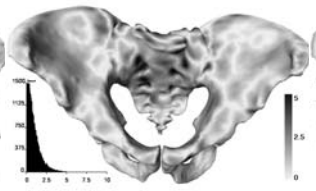


Fig. 5. $\mathcal{S}_T$ for C^1 wind

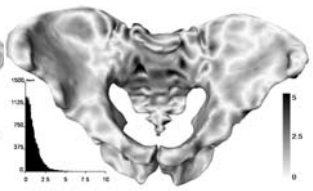


Fig. 6. $\mathcal{S}_T$ for C^∞ wind

A quantitative comparison between all different wind discretizations is issue of Table 2. Therein the column DOFs denote the number of freely choosable vectors $\alpha_{j,t}$ for fixed t. All methods provide acceptable matches with respect to the fixed level of detail $\sigma_0 = 8$. Especially the surface $\mathcal{S}_T$ corresponding to C^1 wind is also visually closest to $\mathcal{T}$ although it is obtained via less wind DOFs compared to the approach from Exoshape. This fact stresses the potential of sub-optimal, albeit simpler wind parameterizations and decoupling the discretization of the spaces W_m from V.

Table 2. One-sided surface distances between $\mathcal{S}_T$ and $\mathcal{T}$ for $m = 0$

v_t	DOFs	mean	stddev	rms	max
C^0	1·(568-133) = 435	1.18	1.14	1.64	11.33
C^1	8·120 = 960	0.95	0.91	1.32	9.16
C^∞	1746	0.97	0.93	1.34	9.70
no wind, $\mathrm{dist}_{x\in\mathcal{S}}(x, \mathcal{T})$		5.08	3.87	6.39	21.05

Appendix

In the following we prove Theorem 2.

[1] http://www-sop.inria.fr/asclepios/projects/Health-e-Child/ShapeAnalysis/

Proof. First we consider the variation of the kinetic energy, i.e. $\omega_m = 0$ for all m. One directly calculates

$$(\nabla \hat{J})_t = 2\gamma \mathbf{S}\boldsymbol{\alpha}_t \ . \tag{8}$$

Let us now consider the contrary case, i.e. $\gamma = 0$. We aim to compute $\nabla_{\boldsymbol{\alpha}} E^m$ for some fixed m. Variation of $E = E^m$ w.r.t. $\boldsymbol{\alpha}_t$ in direction $\tilde{\boldsymbol{\alpha}}_t$ gives

$$\tilde{E}^m = (\partial_{\mathbf{x}_T} E)\tilde{\mathbf{x}}_T + (\partial_{\mathbf{a}_T} E)\tilde{\mathbf{a}}_T \ . \tag{9}$$

There holds

$$\tilde{\mathbf{x}}_t = \int_0^t \tilde{v}_s(\mathbf{x}_s)\, ds = \int_0^t \boldsymbol{\varphi}_s \tilde{\boldsymbol{\alpha}}_s \, ds \ . \tag{10}$$

From Theorem 1 the evolution of m-current attributes can be written as

$$\mathbf{a}_t' = \mathbf{g}(\boldsymbol{\alpha}_t, \mathbf{x}_t, \mathbf{a}_t) = \mathbf{g}_t \quad \text{with} \quad \mathbf{a}(0) = \mathbf{a}_0 \ . \tag{11}$$

Its variation in direction $\tilde{\boldsymbol{\alpha}}_t$ satisfies

$$\tilde{\mathbf{a}}_t' = (\partial_{\boldsymbol{\alpha}} \mathbf{g}_t)\tilde{\boldsymbol{\alpha}}_t + (\partial_{\mathbf{x}} \mathbf{g}_t)\tilde{\mathbf{x}}_t + (\partial_{\mathbf{a}} \mathbf{g}_t)\tilde{\mathbf{a}}_t \quad \text{with} \quad \tilde{\mathbf{a}}(0) = 0 \ .$$

It remains to express $\tilde{\mathbf{a}}_t$. We therefore introduce the flow $\frac{dF_{st}}{dt} = (\partial_{\mathbf{a}} \mathbf{g}_t) F_{st}$ with $F_{tt} = I$ and get

$$\begin{aligned}
\tilde{\mathbf{a}}_t &= \int_0^t F_{ut}((\partial_{\boldsymbol{\alpha}} \mathbf{g}_u)\tilde{\boldsymbol{\alpha}}_u + (\partial_{\mathbf{x}} \mathbf{g}_u)\tilde{\mathbf{x}}_u)\, du \\
&= \int_0^t F_{ut}(\partial_{\boldsymbol{\alpha}} \mathbf{g}_u)\tilde{\boldsymbol{\alpha}}_u \, du + \int_0^t \int_0^u F_{ut}(\partial_{\mathbf{x}} \mathbf{g}_u)\boldsymbol{\varphi}_s \tilde{\boldsymbol{\alpha}}_s \, ds \, du \\
&= \int_0^t \left(F_{ut}(\partial_{\boldsymbol{\alpha}} \mathbf{g}_u) + \int_u^t F_{st}(\partial_{\mathbf{x}} \mathbf{g}_s)\, ds\, \boldsymbol{\varphi}_u \right) \tilde{\boldsymbol{\alpha}}_u \, du \ .
\end{aligned}$$

In particular there holds

$$\tilde{\mathbf{a}}_T = \int_0^T \left(F_{tT}(\partial_{\boldsymbol{\alpha}} \mathbf{g}_t) + \int_t^T F_{sT}(\partial_{\mathbf{x}} \mathbf{g}_s)\, ds\, \boldsymbol{\varphi}_t \right) \tilde{\boldsymbol{\alpha}}_t \, dt \ . \tag{12}$$

Combining (9), (10) and (12) we have

$$\begin{aligned}
\tilde{E}^m &= \int_0^T (\partial_{\mathbf{x}_T} E)\boldsymbol{\varphi}_t \tilde{\boldsymbol{\alpha}}_t + (\partial_{\mathbf{a}_T} E) \left(F_{tT}(\partial_{\boldsymbol{\alpha}} \mathbf{g}_t) + \int_t^T F_{sT}(\partial_{\mathbf{x}} \mathbf{g}_s)\, ds\, \boldsymbol{\varphi}_t \right) \tilde{\boldsymbol{\alpha}}_t \, dt \\
&= \int_0^T \Bigg(\underbrace{\left[(\partial_{\mathbf{x}_T} E) + \int_t^T (\partial_{\mathbf{a}_T} E) F_{sT}(\partial_{\mathbf{x}} \mathbf{g}_s)\, ds \right]}_{=:\boldsymbol{\eta}_t^*} \boldsymbol{\varphi}_t + \underbrace{(\partial_{\mathbf{a}_T} E) F_{tT}}_{=:\boldsymbol{\zeta}_t^*} (\partial_{\boldsymbol{\alpha}} \mathbf{g}_t) \Bigg) \tilde{\boldsymbol{\alpha}}_t \, dt \\
&= \int_0^T \left(\boldsymbol{\eta}_t^* \boldsymbol{\varphi}_t + \boldsymbol{\zeta}_t^* (\partial_{\boldsymbol{\alpha}} \mathbf{g}_t) \right) \tilde{\boldsymbol{\alpha}}_t \, dt \ .
\end{aligned} \tag{13}$$

Since $F_{st}F_{ts} = I$ and $\frac{dF^*_{st}}{ds} = -(\partial_{\mathbf{a}}\mathbf{g}_s)^* F^*_{st}$ we have in particular the integral form $F^*_{ts} = I + \int_t^s (\partial_{\mathbf{a}}\mathbf{g}_u)^* F^*_{us}\, du$. This helps to simplify

$$\begin{aligned}\boldsymbol{\zeta}_t = F^*_{tT}(\nabla_{\mathbf{a}_T} E) &= \Big(I + \int_t^T (\partial_{\mathbf{a}}\mathbf{g}_s)^* F^*_{sT}\, ds\Big)(\nabla_{\mathbf{a}_T} E) \\ &= \nabla_{\mathbf{a}_T} E + \int_t^T (\partial_{\mathbf{a}}\mathbf{g}_s)^* F^*_{sT}(\nabla_{\mathbf{a}_T} E)\, ds = \nabla_{\mathbf{a}_T} E + \int_t^T (\partial_{\mathbf{a}}\mathbf{g}_s)^* \boldsymbol{\zeta}_s\, ds \qquad (14)\end{aligned}$$

$$\boldsymbol{\eta}_t = \nabla_{\mathbf{x}_T} E + \int_t^T (\partial_{\mathbf{x}}\mathbf{g}_s)^* F^*_{sT}(\nabla_{\mathbf{a}_T} E)\, ds = \nabla_{\mathbf{x}_T} E + \int_t^T (\partial_{\mathbf{x}}\mathbf{g}_s)^* \boldsymbol{\zeta}_s\, ds\ . \qquad (15)$$

Collecting (8), (13), (14) and (15) yields the assertion. □

Acknowledgement. We thank Stanley Durrleman from the University of Utah for the fruitful discussion and helpful suggestions at the initial phase of this paper. We further acknowledge support by DFG-MATHEON Project F2. Moreover we thank the referees for their comments which helped to improve the quality of this paper.

References

1. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV 61(2), 139–157 (2005)
2. Cao, Y., Miller, M.I., Winslow, R.L., Younes, L.: Large deformation diffeomorphic metric mapping of vector fields. IEEE Trans. Med. Imag. 24(9), 1216–1230 (2005)
3. Cotter, C.: The variational particle-mesh method for matching curves. Journal of Physics A: Mathematical and Theoretical 41(34), 344003 (2008)
4. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. Q. Appl. Math. 56(3), 587–600 (1998)
5. Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Phd, Univ. Nice (2010)
6. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Statistical models of sets of curves and surfaces based on currents. MedIA 13(5), 793–808 (2009)
7. Federer, H.: Geometric measure theory. Repr. of the 1969 ed. Classics in Mathematics. Berlin: Springer-Verlag (1996)
8. Glaunès, J., Qiu, A., Miller, M.I., Younes, L.: Large deformation diffeomorphic metric curve mapping. IJCV 80(3), 317–336 (2008)
9. Glaunès, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. In: CVPR. Los Alamitos: IEEE Comput. Soc. vol. 2, pp. 712–718 (2004)
10. Glaunès, J.: Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique. Ph.D. thesis, Université Paris 13 (2005)
11. Kirk, B.S., Peterson, J.W., Stogner, R.H., Carey, G.F.: `libMesh`: A C++ Library for Parallel Adaptive Mesh Refinement/Coarsening Simulations. Engineering with Computers 22(3–4), 237–254 (2006)
12. Marsland, S., Twining, C.: Constructing diffeomorphic representations for the groupwise analysis of non-rigid registrations of medical images. IEEE Trans. Med. Imag. 23(8), 1006–1020 (2004)

13. Mattheij, R.M., Molenaar, J.: Ordinary differential equations in theory and practice. Reprint of the 1996 original. Classics in Applied Mathematics. SIAM (2002)
14. Morgan, F.: Geometric measure theory. A beginner's guide. 4th ed. Elsevier (2009)
15. Morita, S.: Geometry of differential forms. American Mathematical Society (2001)
16. Risser, L., Vialard, F.X., Wolz, R., Holm, D.D., Rueckert, D.: Simultaneous fine and coarse diffeomorphic registration: Application to atrophy measurement in alzheimer's disease. In: MICCAI'10. vol. 13, pp. 610–617 (2010)
17. Sommer, S.H., Nielsen, M., Lauze, F.B., Pennec, X.: A multi-scale kernel bundle for LDDMM: Towards sparse deformation description across space and scales. In: IPMI. LNCS, Springer (2011)
18. Trouvé, A.: An infinite dimensional group approach for physics based models in pattern recognition. Tech. rep., Johns Hopkins University (1995)
19. Vaillant, M., Glaunès, J.: Surface matching via currents. In: Christensen, G.E., Sonka, M. (eds.) IPMI, LNCS, vol. 3565, pp. 381–392. Springer (2005)
20. Vaillant, M., Miller, M.I., Younes, L., Trouvé, A.: Statistics on diffeomorphisms via tangent space representations. NeuroImage 23(Supplement 1), 161–169 (2004)
21. Younes, L.: Shapes and diffeomorphisms. App. Math. Sc. 171. Springer (2010)

Frame to Frame Diffeomorphic Motion Analysis from Echocardiographic Sequences

Zhijun Zhang[1], David J. Sahn[1,2], and Xubo Song[1]

[1] Department of Biomedical Engineering
[2] Department of Pediatric Cardiology
Oregon Health and Science University
20000 NW Walker Road, Beaverton, OR 97006, USA
{zhangzhi,songx,sahn}@ohsu.edu

Abstract. Quantitative motion analysis from echocardiography is an important yet challenging problem. We develop a motion estimation algorithm for echocardiographic image sequences based on diffeomorphic image registration in which the velocity field is spatiotemporally smooth. The novelty of this work is that instead of optimizing a functional of velocity field which consists of similarity metrics between a reference image to each of the following images (*first-to-follow*), we optimize a functional which is a sum of similarity metrics of each two consecutive images (*frame-to-frame*). This method can reduce the bias effect of using a single image as reference. It also improves registration accuracy since consecutive frames usually have higher dependency than frames far away. We validate our method by using both simulated images with known ground truth and *in vivo* pig heart images with sonomicrometry. Tests indicate that our frame-to-frame motion estimation method is more accurate than first-to-follow method.

1 Introduction

Quantitative analysis of cardiac deformation and motion is important for studying heart function. Many illnesses related to ischemia or infarct can be recognized from the motion and deformation abnormalities [1]. Techniques to discriminate the abnormal motion and accurately locate regions with motion abnormality are critical to identify the disease and to evaluate the treatment. Echocardiography (echo) is the most widely used image modality because it is non-ionizing, real-time, cost-effective and convenient. With the development of the new transducer array technology, 3D echo can now provide real-time images of the whole heart [2]. However, due to the low signal-noise-ratio, general methods for motion estimation do not work well on echo images. In addition, the 4D (3D+t) data is acquired with a compromise that both the spatial and temporal resolutions are reduced comparing to 2D+t sequences. As a result, 3D motion analysis from echo sequences remains a challenging problem.

Cardiac motion analysis algorithms can be classified into three categories: model-based, feature-based and voxel-based methods. Lots of cardiac models

have been proposed for motion analysis and segmentations [3]. Surface models such as super-quadrics are used for motion analysis by fitting the model with a sequence of images [4, 5], however, these models only estimate the deformation on the model surfaces. Comaniciu *et al.* [6] proposed a Kalman filter based shape tracking method by using information fusion framework with a probabilistic subspace model constraint. In this work, a shape model needs to be learned and the motion estimation is limited along the contour points. Volumetric models such as dynamic finite elements have been used to estimate the deformation inside the myocardium [7, 8]. Wang *et al.* [9] tracked myocardial surface points by maximizing the likelihood of a combined surface and a two-steps motion prediction model. Both the initial myocardial surface detector and the motion prediction model need to be learned in advance. Generally speaking, deformable model based methods needs prior knowledge related to the models and their generation needs some sort of human interaction. Feature-based methods use landmarks such as the tagged lines to fit the deformable model such as 4D B-spline [10]. However, in echocardiography due to lack of stable landmarks in the myocardium and artificial features such as the tagged line are not available, feature-based method is difficult to estimate the deformation by using trustable correspondence. Voxel-based methods require no manual intervention, they estimate spatially dense transformations from all image voxels directly. This method can be implemented as an automatic method and we focus our work on this approach. Voxel based image registration methods such as optical-flow [11] and B-spline based methods [12, 13] have been proposed for cardiac motion analysis from echo images. However, the motion analysis problem is simplified into a series of independent pairwise image registrations and the temporal motion smoothness is not considered. To enforce the temporal consistency of particle motion, many temporal models have been used. Carbayo *et al.* [14] proposed a spatiotemporal deformation model for cardiac motion tracking. A 2D+t B-spline transformation with spatiotemporal smoothness is used with the first frame as the reference. A 3D+t extension is proposed by Metz *et al.* [15] and the average image is used as the reference. Particle trajectory constraint such as polynomial modeling has been used to regularize the spatiotemporal motion smoothness [16]. Diffeomorphic image registration is a method which the transformation is implicitly spatiotemporally smooth. The transformation between two images is defined as the end point of a velocity field flow which will be obtained by optimization of an energy functional of it. It has a very useful characteristic in computational anatomy that the transformation is one-to-one mapping and topology preserving [17]. This large deformation topology preserving property is preferred in cardiac motion analysis because of the fact that the deformation from the reference frame (usually the end of diastolic) to the mostly contracted frame (the end of systolic) is so large that deformation models without topology preserving constraint may cause the transformation to fold over or tear apart, which is not physically plausible. Beg *et al.* [18] proposed a large deformation diffeomorphic metric mapping algorithm (LDDMM) in which the smooth velocity field is estimated by optimizing a sum of squared difference (SSD) energy.

Khan *et al.* [19] extended the LDDMM method to analyze the anatomical shape evolution in an image sequence. De Craene *et al.* [20] proposed a method in which the velocity field is defined as temporal piecewise continuous 3D B-spline functions and the B-spline control parameters are estimated by optimization of a parameterized energy function. In a following work [21], the velocity field is defined as a 3D+t spatiotemporal B-spline model to reduce B-spline control points in temporal direction. In both methods, the optimal velocity field minimizes the summed dissimilarity metrics between the first frame and each of the unwarped subsequent frames, which we call them *first-to-follow* methods. It has been presented theoretically and experimentally that speckle pattern will change under large deformation [22] and that registration of frames further away from the reference is less accurate due to speckle de-correlation [12]. We propose a diffeomorphic registration method with a spatiotemporally smooth velocity field which minimizes the summed SSDs of the unwarped consecutive frames (*frame-to-frame* method). Our registration method is tested with simulated and *in-vivo* pig datasets, the results show that the accuracy is improved over first-to-follow method.

2 Method

2.1 Diffeomorphic Image Sequence Registration

A diffeomorphism flow is a dynamic system with each of the diffeomorphism to be a state in a differentiable manifold [23]. We define a flow $\boldsymbol{\phi}(\mathbf{x},t), t \in [0,T], \mathbf{x} \in \Omega \subset R^d (d=2,3)$ with its smooth velocity field $\boldsymbol{v}(\mathbf{x},t)$ by using the differential equation of $\frac{d\boldsymbol{\phi}}{dt} = \boldsymbol{v}(\boldsymbol{\phi}(\mathbf{x},t),t)$. It has been proven in [24] that if $\boldsymbol{v}(\mathbf{x},t)$ is smooth enough with a differential operator L in a Sobolev space V, then the transformation $\boldsymbol{\phi}(\mathbf{x},t)$ will be a group of diffeomorphisms with t varying from 0 to T. The diffeomorphic image registration is stated as a variational problem, that given two images I_0 and I_1, to find an optimal velocity field $\hat{\boldsymbol{v}}$ which minimizes an energy functional consisting of a sum of squared difference (SSD) and a geodesic distance metric between $\boldsymbol{\phi}(\mathbf{x},0)$ and $\boldsymbol{\phi}(\mathbf{x},T)$ [18]:

$$\hat{\boldsymbol{v}} = arg \inf_{\boldsymbol{v}\in V} \lambda \int_0^T ||\boldsymbol{v}(\mathbf{x},t)||_V^2 dt + \int (I_0(\mathbf{x}) - I_1(\boldsymbol{\phi}(\mathbf{x},T)))^2 d\mathbf{x}, \tag{1}$$

with λ being the weight to balance these two energies. If we have a sequence of N_f images to be registered, the similarity metric consists of the SSD of the difference between a reference frame I_0 and each of the deformed subsequent frames $I_k(\boldsymbol{\phi}(\mathbf{x},t_k)), k = 1,2,...,N_f$. Then the mathematical form can be in a similar form as [19]:

$$\hat{\boldsymbol{v}} = arg \inf_{\boldsymbol{v}\in V} \lambda \int_0^T ||\boldsymbol{v}(\mathbf{x},t)||_V^2 dt + \sum_{k=1}^{k=N_f} \int (I_0(\mathbf{x}) - I_k(\boldsymbol{\phi}(\mathbf{x},t_k)))^2 d\mathbf{x}. \tag{2}$$

This scheme has two disadvantages: first, the speckle de-correlation between far away frames are high which may cause correspondence ambiguity between

the two images; second, it take longer time to converge since the difference between reference frame to far away frames is bigger than that of the consecutive frames. Instead of optimizing each deformed subsequent frames to be similar to the reference frame, we propose a variational energy which minimizes the difference between every two deformed consecutive frames $I_{k-1}(\boldsymbol{\phi}(\mathbf{x}, t_{k-1}))$ and $I_k(\boldsymbol{\phi}(\mathbf{x}, t_k))$:

$$\hat{\boldsymbol{v}} = arg \inf_{\boldsymbol{v} \in V} \lambda \int_0^T ||\boldsymbol{v}(\mathbf{x}, t)||_V^2 dt + \sum_{k=1}^{k=N_f} \int (I_{k-1}(\boldsymbol{\phi}(\mathbf{x}, t_{k-1})) - I_k(\boldsymbol{\phi}(\mathbf{x}, t_k)))^2 d\mathbf{x}, \tag{3}$$

we denote the two energy terms in Eqn.(3) as E_{reg} and E_{ssd} respectively. This method will find a flow of diffeomorphisms which have the shortest geodesic path in the manifold and simultaneously minimizes summed errors of each two neighboring images. This will improve the accuracy of the transformation since the neighboring frames generally have higher intensity correlation than those which are not neighbors. It can also reduce the chance that the transformation is biased due to noise in the reference images. Our frame-to-frame method is different from the method which simply estimates the diffeomorphisms between consecutive frames and then composites together. In our method, minimizing the difference between two consecutive frames jointly optimizes all the velocity field before and between the time of these two frames.

The direct solution for this variational framework is expensive. Alternatively, a parameterized representation of the velocity field is used [25], where the velocity field is represented as a series of B-spline functions and the displacement field can be expressed as the forward Euler integral of velocity field. We use a spatiotemporally smooth B-spline function to represent the velocity field. It is defined as $\boldsymbol{v}(\mathbf{x}, t_k) = \sum \mathbf{c}_{i;k} \boldsymbol{\beta}(\mathbf{x} - \mathbf{x}_i)$, with $\mathbf{c}_{i;k}$ being the B-spline control vectors at t_k located on a uniform grid of $\mathbf{x}_i$, $\boldsymbol{\beta}(\mathbf{x} - \mathbf{x}_i)$ is the 3D B-spline kernel function which is the tensor product of the 1-D B-spline functions. Define $\boldsymbol{\phi}_k = \boldsymbol{\phi}(\mathbf{x}, t_k)$ the transformation at time step t_k, we assume the velocity is piecewise constant within a time step, then we have $\boldsymbol{\phi}_k = \boldsymbol{\phi}_{k-1} + \boldsymbol{v}(\boldsymbol{\phi}_{k-1}, t_{k-1}) \Delta t$, with $\boldsymbol{\phi}_0(\mathbf{x}) = \mathbf{x}$, $k = 1, 2, ..., N_t$, with N_t being the total number of time steps of the velocity field. Without loss of generality, we can have $\Delta t = 1$. In our test, we use one time step in velocity field between two neighboring frames since the deformation between them is usually small, that is $N_t = N_f$. However, our method easily generalizes to multiple time steps between frames if the deformation between two consecutive frames is large. The energy functional will be a parameterized function of $\mathbf{c}_{i;k}$ and it can be optimized by using a steepest descent method.

2.2 Regularization

In order to assure the $\boldsymbol{\phi}(\mathbf{x}, t)$ to be diffeomorphic, we need to define $\boldsymbol{v}(\mathbf{x}, t)$ to be spatiotemporally smooth under a differential operator L. The linear operator we choose is: $L = \nabla^2 \boldsymbol{v} + w_t \frac{d\boldsymbol{v}}{dt}$, with $\nabla^2(\cdot)$ being a Laplacian operator and w_t a

constant weight. In the discrete time form of velocity field, the time integral of the norm in V space of Eqn.(3) will be: $E_{reg} = \sum_{k=1}^{N_t}\sum_{\mathbf{x}}(\nabla^2 \boldsymbol{v}_k)^2 + w_t \sum_{k=2}^{N_t}\sum_{\mathbf{x}}|\boldsymbol{v}_k(\mathbf{x} + \boldsymbol{v}_{k-1}\Delta t) - \boldsymbol{v}_{k-1}|^2$, with $\boldsymbol{v}_k = \boldsymbol{v}(\mathbf{x}, k)$. The first term makes the velocity field spatially smooth which is denoted as E_{sr}. The second term keeps the particle velocity smooth and it is denoted as E_{tr}. The overall effect is to keep the velocity field spatiotemporally smooth.

2.3 Optimization

We use a steepest descent method to optimize the parameterized function. The derivative of the total registration energy with respect to the transformation parameters will be calculated analytically. The derivative of the similarity metric with respect to the B-spline parameters $\mathbf{c}_{i;k'}$ is:

$$\frac{\partial E_{ssd}}{\partial \mathbf{c}_{i;k'}} = \sum_{k=1}^{N_f}(I_k(\boldsymbol{\phi}_k) - I_{k+1}(\boldsymbol{\phi}_{k+1}))(\nabla I_k(\boldsymbol{\phi}_k)\frac{\partial \boldsymbol{\phi}_k}{\partial \mathbf{c}_{i;k'}} - \nabla I_{k+1}(\boldsymbol{\phi}_{k+1})\frac{\partial \boldsymbol{\phi}_{k+1}}{\partial \mathbf{c}_{i;k'}}), \quad (4)$$

with $\frac{\partial \boldsymbol{\phi}_k}{\partial \mathbf{c}_{i;k'}}$ being the Jacobian matrix of transformation at time step k with respect to the ith B-spline coefficient in k'th frame. It can be calculated with chain rule and it is zero when $k' \geqslant k$. For detailed computation refer to [21].

For the derivative of the spatial and temporal regularization energies with respect to the mth component of $\mathbf{c}_{i;k}$, we have:

$$\frac{\partial E_{sr}}{\partial c_{i,m;k}} = \sum_{\mathbf{x}\in\Omega'} \boldsymbol{\beta}_m^{''}(\mathbf{x} - \mathbf{x}_i), \quad (5)$$

with Ω' being the local support of the B-spline kernel function, and $\boldsymbol{\beta}_m^{''}(\cdot)$ being the second derivative of the B-spline function with respect to mth component. Considering that the displacement between two time step is small, we have:

$$\frac{\partial E_{tr}}{\partial c_{i,m;k}} \approx w_t \sum_{\mathbf{x}\in\Omega'} (2 * v_{i,m;k} - v_{i,m;k-1} - v_{i,m;k+1})\boldsymbol{\beta}(\mathbf{x} - \mathbf{x}_i). \quad (6)$$

The registration energy can be optimized by starting from initial position and descending along the negative gradient direction at each iteration until there is no significant decrease.

2.4 Implementation

In our implementation, we use a series of B-spline transformations with grid spacing of 10 in each dimension to represent the velocity field. The values of λ and w_t are set to be 0.1 and 0.5. The algorithm is implemented with Matlab under a windows XP 64 bit system on a machine with 2.13GHz Xeon 8 cores CPU and 6GB memory. It takes about 1 hour to register a 3D (3 minutes for a 2D) sequence with 20 frames.

3 Experiment and Data

We use both simulated and real data to validate our algorithm. In the simulated data experiment, a longitudinal view of a diastolic left ventricle (LV) image with size of 274×192 is used as the reference image. This frame is then deformed with a series of continuous displacement field functions. The deformations are symmetrical along the long axis of the LV to simulate the myocardial contraction effect along radial and longitudinal directions. The displacement functions are in form of: $f_x(i) = a_x sin\frac{\pi(x-x_c)}{2r_d} sin(\frac{i\pi}{N_f})$ and $f_y(i) = a_y sin\frac{\pi(y-y_{apex})}{2(y_{base}-y_{apex})}(sin(\frac{i\pi}{N_f} + \frac{\pi}{16}) - sin\frac{\pi}{16})$, with x_c, r_d the axis center coordinate and the average axial radius of LV, y_{apex} and y_{base} the height of base and apex planes, N_f and i the number of frames and the frame index, and a_x, a_y are the magnitudes of displacement fields. An image sequence with $N_f + 1$ frames is generated when i varies from 0 to N_f to simulate the cardiac motion in one cycle.

We carry out two experiments for the simulated data. In the first experiment, three sequences with 20 frames each are simulated with multiplicative speckle noise of variance 0.06, 0.08 and 0.10 added. The reference frame and the 10th frame with speckle noise variance 0.10 are shown together with the ground truth displacement field in Fig.1. In the second simulated experiment, we first generate 20 frames without deformation by adding independent speckle noise of variance 0.10 to the reference frame. Then each of the frame I_i will have a percentage p pixels replaced with the intensity at the same position in frame I_{i-1}. By updating noisy image one by one we assure that the two consecutive frames to have noise overlap ratio of p. Each frames will then be deformed by using the ground truth displacement fields. We simulate two sequences with overlap ratio of 0.2 and 0.4 respectively.

Fig. 1. The reference frame and the $10th$ frame in speckle variance 0.1 test and the displacement field (only displacement field inside a bell-shaped mask is displayed).

A real world dataset is acquired from an open-chest pig by using a Philips IE33 system. For validation, we installed six sonomicrometers in the heart wall. The distances between each pair of the sonomicrometers are recorded with the image sequences and are used as ground truth to compare with the tracked distance in the echo images. The images are resampled into volume sequences of $160 \times 100 \times 128$ with voxel size $1mm \times 1mm \times 1mm$. The crystal coordinates in the reference frame are manually denoted.

4 Result

In the first simulated experiment, we compare our frame-to-frame method with the first-to-follow method by tracking the trajectories of the points in the myocardial wall during the motion process. The estimated trajectories of four example points from a small region of the myocardium for speckle noise variance 0.06 test are shown in Fig.2, where the ground truth trajectories are overlaid for comparison. We can see generally coordinates of the points in each time step in our method are closer to the ground truth position than the first-to-follow method.

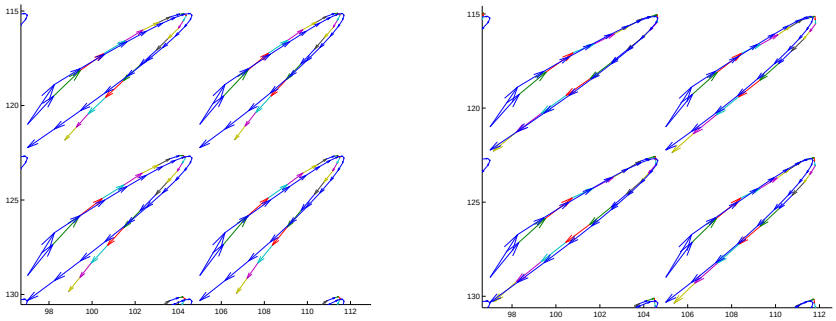


Fig. 2. Points trajectories in the first-to-follow method (left) and the frame-to-frame method (right). The ground truth trajectories (blue) are overlaid with the estimated curves (multiple color) for comparison. The arrow shows the velocity at each time step.

In Fig.3 we illustrate the motion estimation errors in both x and y coordinates in noise level 0.08 dataset. We can see that the motion estimation errors of x and y coordinates in most of frames in our method are closer to zero than those of the first-to-follow method. The figure also shows that our method has a smaller error variance in both coordinates. The mean of magnitude of errors in the three noise variance levels in frame-to-frame method are 0.23, 0.26 and 0.32, while in first-to-follow method they are 0.38, 0.45 and 0.56 respectively.

In the second test, the motion estimation errors in y coordinates for two methods are shown in Fig.4. The results are similar for x coordinate errors. We can see the results of frame-to-frame method are better than those in first-to-follow method in means and standard deviations of errors. We can see for the frame-to-frame method, when the intensity correlation between two images is increased, the mean and variance of the registration error are decreased. In the results of first-to-follow methods, the error mean does not change obviously when the correlation between consecutive frames are increased.

In the *in-vivo* open-chest pig test, we compare the performance of the two algorithms by computing the correlations over time between the algorithm-derived pair-wise distances with sonomicrometry, shown in Table.1. Sonomicrometry pro-

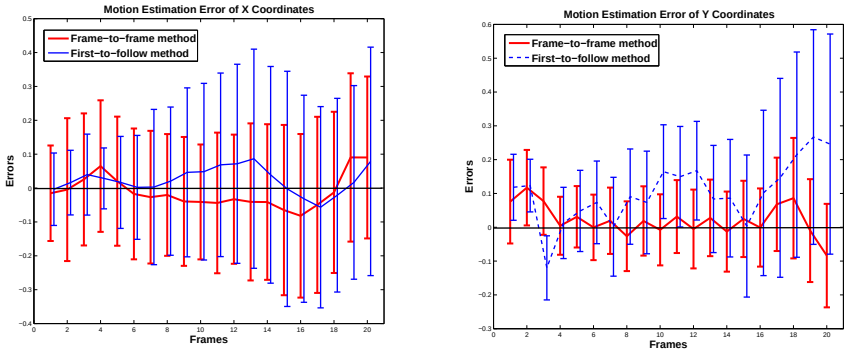


Fig. 3. The motion estimation errors in x (left) and y (right) coordinates of frame-to-frame method and first-to-follow method. The error bars shows the standard deviation of the errors in each frames. The horizontal lines represent the curves of zero mean transformation errors to help comparison.

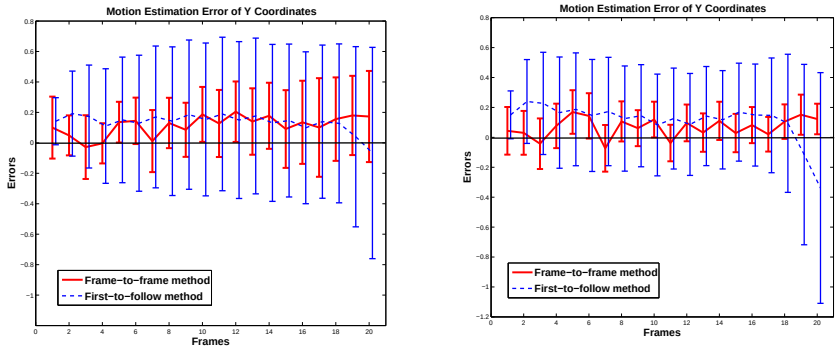


Fig. 4. The motion estimation errors in y coordinates of frame-to-frame method and first-to-follow method. The left and right figures show the results of 20 and 40 percent noise overlap tests. The horizontal lines show the zeros mean transformation errors.

vide the ground truth distances between each two of the crystals. We can clearly see the improvement of our proposed method.

Table 1. The correlations between the estimated pair-wise distances and those from the sonomicrometry, with frame-to-frame method (numbers to the left) and the first-to-follow method (numbers to the right). Numbers 1-6 index the six sonomicrometry markers.

	1	2	3	4	5	6
1	1.0/1.0	0.936/0.907	0.901/0.885	0.913/0.902	0.948/0.923	0.859/0.831
2	0.936/0.907	1.0/1.0	0.881/0.856	0.927/0.904	0.887/0.838	0.951/0.916
3	0.901/0.885	0.881/0.856	1.0/1.0	0.825/0.786	0.902/0.905	0.819/0.788
4	0.913/0.902	0.927/0.904	0.825/0.786	1.0/1.0	0.937/0.919	0.934/0.902
5	0.948/0.923	0.887/0.838	0.902/0.905	0.937/0.919	1.0/1.0	0.918/0.873
6	0.859/0.831	0.951/0.916	0.819/0.788	0.934/0.902	0.918/0.873	1.0/1.0

5 Conclusion

We propose a large deformation diffeomorphic registration method by minimizing the difference between every *consecutive images*. Simulation test shows that our frame-to-frame method has a higher estimation accuracy of motion than the first-to-follow method. Validation with sonomicrometry also shows that our motion estimation result has higher consistency with real data.

Acknowledgement

This paper is supported by a NIH/NHLBI grant 1R01HL102407-01 awarded to Xubo Song and David Sahn.

References

1. Fukuda, K., Oki, T., Tabata, T., Iuchi, A., Ito, S.: Regional left ventricular wall motion abnormalities in myocardial infarction and mitral annular descent velocities studied with pulsed tissue Doppler imaging. JASE, 11(9), 841–848 (1998)
2. Hung, J., Lang, R., Flachskampf, F., Shernan, S.K., McCulloch, M.L., Adams, D.B., Thomas, J., Vannan, M., Ryan, T.: 3D Echocardiograhy: A Review of the Current Status and Future Directions. JASE, 20(3), 213–233 (2007)
3. A.F. Frangi, W.J. Niessen, and Max A. Viergever. "Three-Dimensional Modeling for Functional Analysis of Cardiac Images: A Review." IEEE Trans. Med. Imaging, vol.20, no.1, pp.1–25, 2001.
4. Eric Bardinet, Laurent D. Cohen and Nicholas Ayache. "Tracking and motion analysis of the left ventricle with deformable superquadrics. Medical Imaage analysis, vol.1, no.2, pp:129–149, 1996.
5. Jinah Park, Dimitri Metaxas, Alistair A. Young and Leon Axel. "Deformable Models with Parameter Functions for Cardiac Motion Analysis from Tagged MRI Data." IEEE Transactions on Medical Imaging, vol.15, no.3, pp.278–289, June, 1996.
6. Comaniciu, D, Zhou, X., Krishnan, S.: Robust tracking of myocardial border: An information fusion approach. IEEE Trans. Med. Imag. 23(7), 849–860 (2004)

7. Papademetris, X., Sinusas, A.J., Dione, D.P., Duncan, J.S.: Estimation of 3D left ventricular deformation from echocardiography. Med. Imag. Anal. 5(1), 17–28 (2001)
8. Joel Schaerera, Christopher Casta, Jerome Pousina and Patrick Claryssea. "A dynamic elastic model for segmentation and tracking of the heart in MR image sequences." Medical Image Analysis, vol.14, pp.738–749, 2010.
9. Wang, Y., Georgescu, B., Houle, H., Comaniciu,D.: Volumetric Myocardial Mechanics from 3D+t Ultrasound Data with Multi-model Tracking. In STACOM 2010. LNCS vol. 6364, 184–193. Springer, Heidelberg (2010)
10. J. Huang, D. Abendschein, V. Davila, A.A. Amini, "Four-dimensional LV tissue tracking from tagged MRI with a 4D B-spline model." IEEE Trans Medical Imaging, vol.18, no.10, 1999.
11. Suhling, M., Arigovindan, M., Jansen, C., Hunziker, P., Unser, M.: Myocardial Motion Analysis from B-Mode Echocardiograms. IEEE Trans. Image Process., 14(4) 525–536 (2005)
12. Elen, A., Choi, H.F., Loeckx, D., Gaom, H., Claus, P., Suetens, P., Maes, F., D'hooge, J.: Three-dimensional cardiac strain estimation using spatio-temporal elastic registration of ultrasound images: a feasibility study. IEEE Trans. Med. Imag., 27(11), 1580–1591, (2008)
13. Myronenko, A., Song, X.B., David, J.S: LV Motion Tracking from 3D Echocardiography Using Textural and Structural Information. MICCAI 2007. 428–435 (2007)
14. Ledesma-Carbayo, M.J., Mah-Casado, P., Santos, A., Prez-David, E., GarMA, Desco, M.: Spatio-Temporal Nonrigid Registration for Ultrasound Cardiac Motion Estimation. IEEE Trans. Med. Imag., 24(9), 1113–1126 (2005)
15. Metz, C.T., Klein, S., Schaap, M., Walsum, T., Niessen, W.J.: Nonrigid registration of dynamic medical imaging data using nD+t B-splines and a groupwise optimization approach. Med. Imag. Anal. 15(2), 238–249 (2011)
16. Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T.: Four-dimensional deformable image registration using trajectory modeling. Physics in Medicine and Biology, 55, 305–327 (2010)
17. Miller, M.I; Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. Neuroimage, 23(1), 19–33 (2004)
18. Beg, M.F.; Miller, M.I.; Trouve, A.; Younes, L.; Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. INT J COMPUT VISION. 61(2), 139–157 (2005)
19. Khan, A.R.; Beg, M.F.; Representation of time-varying shapes in the large deformation diffeomorphic framework. ISBI 2008, 1521–1524 (2008)
20. Craene, M.; Camara,O.; Bijnens, B.H.; Frangi, A.F.: Large Diffeomorphic FFD Registration for Motion and Strain Quantification from 3D-US Sequences, Functional Imaging and Modeling of the Heart, FIMH'09, LNCS vol.5528, 437–446 (2009)
21. Craene,M.; Piella,G.; Duchateau,N.; Silva,E.; Doltra,A.; Gao,H.; D'hooge,J.; Camara,O.; Brugada,J.; Marta Sitges: Temporal Diffeomorphic Free-Form Deformation for Strain Quantification in 3D-US Images. MICCAI 2010, 1–8 (2010)
22. Meunier, J.: Tissue motion assessment from 3D echographic speckle tracking. Phys. Med. Biol. 43, 1241–1254 (1998)
23. Arrowsmith, D.K.; Place, C. M.: An introduction to dynamical systems. Cambridge Press, UK (1990)
24. Dupuis, P.; Grenander,U.: Variational problems on flows of diffeomorphisms for image matching. Q APPL MATH. 56(3) 587–600 (1998)
25. Ashburner, J.; A fast diffeomorphic image registration algorithm. NeuroImage, 38(1), 95–113 (2007)

An Improved Estimator of GRID Model for Representing Large Diffeomorphic Deformations

Qian Xie and Anuj Srivastava

Department of Statistics, Florida State University, United States
{qxie,anuj}@stat.fsu.edu

Abstract. The growth by random iterated diffeomorphisms (GRID) model seeks to decompose large deformations, caused by growth, anomaly, or anatomical differences, into smaller, biologically-meaningful components. These components are spatially local and parametric, and are characterized by radial deformation patterns around randomly-placed seeds. A sequential composition of these components, using the group structure of diffeomorphism group, models the cumulative deformation. The actual decomposition requires estimation of GRID parameters from observations of large growth, typically from 2D or 3D images. While past papers have estimated parameters under certain simplifying assumptions, including that different components are spatially separated and non-interacting, we address the problem of parameter estimation under the original GRID model that advocates sequential composition of arbitrarily interacting components. Using a gradient-based approach, we present an algorithm for estimation of GRID parameters by minimizing an energy function and demonstrate its superiority over the past additive methods.

Keywords: Large deformation, GRID model, parameter estimation

1 Introduction

The mathematical and statistical modeling of diffeomorphic deformations over time is an important problem with a variety of applications ranging from medical diagnostics to evolutionary biology. The use of medical images, especially the MRI images of human parts, in studying anatomical structures is a growing area of research by itself. Here one uses 2D and/or 3D images taken across time, species, or specimens to compare to extract salient differences in anatomical structures, and to analyze and model their variations both within and across biological classes. These differences may result from standard biological growth, abnormalities, inter-specimen variability, or other reasons. In terms of image-based analysis of anatomical structures, the study of shapes of anatomical parts has become a central idea. For instance, one can use longitudinal image data for tracking biological growth [11, 9, 19, 4, 3] in fetus brains and evaluating tumor growth. A major difficulty in solving such problems is the high dimensionality of image data. The diffeomorphic deformation when estimated from image

sequences can be very high dimensional and not amenable to standard tools from multivariate statistics. Some current methods simplify this analysis by using simplistic measures like lengths, sizes, or areas as indicators of overall shape changes. Some others use relatively simple geometrical models, such as spheres or ellipsoids, to represent shapes in parametric forms and to study the evolution of parameters during growth.

We start with the basic question of how to represent large deformations in a mathematical framework. There is a large body of work on representing differences in imaged objects using deformations of background space [2, 12, 7, 20, 1, 13]. This approach utilizes diffeomorphisms of the underlying coordinate systems to represent and measure shape and other differences. Let an image be $I : [0,1]^d \to \mathbb{R}$, where $d = 1, 2, 3$. A deformation is then a mapping $\Phi : [0,1]^d \to [0,1]^d$, with the resulting deformed image is $I \circ \Phi : [0,1]^d \to \mathbb{R}$. Thus, a point on an anatomical landmark is always observed with the same color intensity; it simply moves to a different location under the deformation. The goal is to use Φ to model, understand and analyze large deformations. These deformations are typically very high dimensional and do not permit standard statistical analysis directly. Therefore, some tools for reducing dimensionality become important. One can apply some standard dimension reduction algorithms, such as PCA, but it is difficult to interpret the resulting representation in biological terms. Durrleman et al. [5] proposed a parametric way of representing large diffeomorphisms by forcing the instantaneous velocity fields to take a parametric form. One starts with a finite number of so-called control points and for each of them specifies a vector that defines the deformation at that point. The vector field over the whole domain is obtained using a Kernel-based interpolation. This approach provides a data-driven sparse parametric method to estimate the large diffeomorphic deformation.

Motivated by the need for biologically-interpretable decompositions of large deformations, Grenander [8] introduced the Growth as Random Iterated Diffeomorphisms (GRID) model. It highlights the role of gene control in biological growths and uses a combination of local, structured deformations to form the large composite deformations. This model has been studied extensively, but mostly from a perspective of synthesis and asymptotics. Some authors proposed a "thermodynamic limit equation" that approximates the growth pattern in a macroscopic way [16, 15]. Portman et al. [14] further developed the GRID model by analyzing the growth patterns at microscopic levels. In addition to synthesis, one is also interested in the inverse problem where we want to decompose large biological growth into smaller biologically-interpretable units. Grenander et al. [18, 8] studied this inverse problem albeit in a limited context. The estimation of growth components was done in two steps: (1) estimate the full deformation between a pair of images that represents biological growth, (2) estimate parameters for growth components under the GRID model, with a major simplification that different components are spatially local and do not interact with each other. With this assumption, the cumulative growth becomes a simple superimposition of different components and one can use standard projection procedures to esti-

mate component parameters. In this paper we seek a solution to the problem of parameter estimation under the original GRID model, without assuming spatial independence of components. This model is different from Durrleman et al. [5] in the sense that it is the individual diffeomorphisms that take the parametric forms, rather than the instantaneous velocity fields. This results in different local deformations around the focal points. For example, in our method the diffeomorphism around a seed is restricted to be radial while in Durrleman et al. [5] there is no such structure.

The estimation follows the two steps as Grenander et al. [8]. In the first step, the full deformation Φ is estimated using the shape matching technique of [10]. As for the second step, since there is a concatenation of deformation associated with different components, the time-ordering of the components becomes important. Due to the nonlinear effect of compositions, it is not possible to solve for GRID parameters using linear methods. This general estimation problem is posed as an optimization problem with a gradient-based minimization of the cost function. The difficulty of getting trapped in local solutions is handled using clever initializations of the gradient algorithm.

2 Grenander's GRID Model

We start by describing the general GRID model as introduced by Grenander [6]. In this model the overall large deformation is modeled as a composition of a sequence of local, elementary deformations. At time t, the elementary growth is a diffeomorphism $\phi_t : [0,1]^d \to [0,1]^d$ such that the point x moves to $\phi_t(x)$. The full deformation is then expressed in the form of the composition of iterated diffeomorphisms.

$$\Phi_{t_1}^{t_n} \equiv \phi_{t_n} \circ \cdots \circ \phi_{t_2} \circ \phi_{t_1} \tag{1}$$

The next step in the GRID model is to simplify diffemorphic components by expressing each ϕ_t in a parametric form. Here the elementary deformation ϕ_t is: (1) assumed to be centered at a point of activation called a *growth seed* x_{seed}, and (2) the growth around the seed is assumed to be radial. Therefore, it is easier to express this local deformation using polar coordinates centered at the seed x_{seed}: $(r,\tau) \longmapsto (\rho(r,\tau),\tau)$ with $r = 0$ denoting the seed. Furthermore, the model assumes that the change in radial distance can be decomposed into two independent parts: $\rho(r,\tau) = r + R(r)\,A(\tau)$. Here, $A : \mathbb{S}^1 \mapsto R$ is called the angular deformation function (ADF) and $R : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is called the radial deformation function (RDF). These individual deformation functions are allowed to take the following forms:

1. The radial deformation function RDF can be one of the following two types:

$$R(r) = \begin{cases} re^{-r^2/c^2}, & r \geq 0,\ c > 0 \\ (r/c)^{p-1}\, e^{-(r/c)}, & r \geq 0,\ p,\ c > 0\ . \end{cases} \tag{2}$$

   In both cases the deformation is zero at the seed ($r = 0$), increases steadily with r, reaches a peak, and then decreases for a further increase in r. The "zone" of influence of a seed is determined by the parameter c.

2. Similarly, the angular deformation function ADF can also take many forms, including:

$$A(\tau) = \begin{cases} a, & \tau \in \mathbb{S}^1,\ a \in \mathbb{R} \\ ae^{\kappa\cos(\tau-\tau_0)-\kappa}, & \tau \in \mathbb{S}^1,\ a \in \mathbb{R},\ \kappa \geq 0,\ \text{and } \tau_0 \in \mathbb{S}^1 \\ \alpha \sin\left(\frac{2\pi\tau}{\sigma}\right), & \tau \in \mathbb{S}^1,\ \alpha \in \mathbb{R},\ \text{and } \sigma > 0. \end{cases} \quad (3)$$

The first case provides an isotropic deformation, the second provides a unimodal deformation with a well-defined growth/decay direction, while the last one provides a sinusoidal variation.

In this paper we will use $R(r) = re^{-r^2/c^2}$ and will study two choices for ADFs: (1) $A(\tau) = a$, and (2) $A(\tau) = ae^{\kappa\cos(\tau-\tau_0)}$. It has been shown that the resulting ϕ_t is a diffeomorphism as long as $-1 < A(\tau) < 2.2408$ [8]. Each such ϕ_t is now characterized by the following set of parameters: $\theta = [\xi, a, c, \kappa, \tau_0] \in \mathbb{R}^6$, where $\xi \in \mathbb{R}^2$ is the seed location.

Problem Statement: Having chosen the model, the estimation problem can be described as follows. Let Φ represent the observed deformation associated with a growth experiment, observed over a time interval $[0, T]$. The goal now is to estimate n, the number of diffeomorphism components and the associated parameters $\theta_j \in \mathbb{R}^6$ for each $j = 1, 2, \ldots, n$.

This problem has been studied by several papers in the past. However, a common simplifying assumption in the past papers is that different seeds are placed away from each other so that there is no or negligible interaction between the corresponding deformations. In this case, the total displacement field $\Psi(x) = \Phi(x) - x$ can be written as a superposition of the displacements resulting from individual seeds:

$$\Psi_{t_1}^{t_n} \equiv \psi_{t_n} + \cdots + \psi_{t_2} + \psi_{t_1}\ , \quad (4)$$

where $\psi_t(x) = \phi_t(x) - x$. This is a very restrictive assumption and reduces the efficacy of the GRID model. The additive model has several problems, including the fact that the set of diffeomorphisms is not a group under the additive model. In the context of biological growth, it is difficult to interpret growth components under the assumption that there is no spatial interaction between them. Also, as shown in Fig. 1, the results of these different models, composite versus additive models are quite different for the same components. It is also illustrated that, for the composite model, the ordering of components is also important in determining the cumulative deformation. In this paper, we study the problem of parameter estimation under the composite model (Eqn. 1) and compare the results with those obtained under the additive model (Eqn. 4).

3 Our Approach

In this section we formulate the problem of parameter estimation as minimization of a certain objective function. The goal is to estimate difeomorphic components

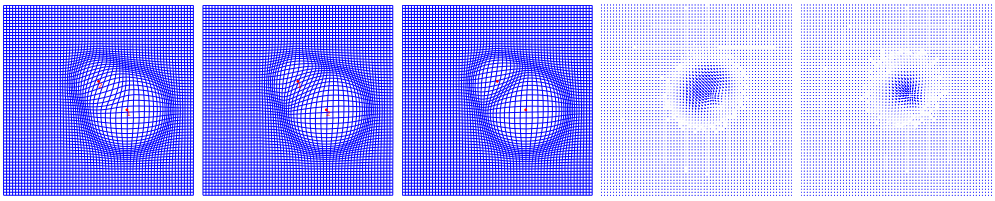


Fig. 1: Cumulative deformation for composite models with: (a) $\Phi = \phi_2 \circ \phi_1$, (b) $\Phi^{'} = \phi_1 \circ \phi_2$, and additive model (c) $\Phi^{''} = x + \psi_1 + \psi_2$, for the same ϕ_1 and ϕ_2. In the remaining panels, we show $\Phi^{'} - \Phi$ (4th panel), and $\Phi^{''} - \Phi$ (5th panel) as vector fields.

$\phi_1,\ \phi_2,\ \ldots, \phi_n$ such that their order composition is as closed to the given Φ as possible. In order words, we can use a distance between Φ and $(\phi_n \circ \cdots \circ \phi_1)$ as the objective function. Although the choice of a geodesic distance in the space of diffeomorphisms (under a suitable Riemannian metric) seems like a natural choice, the use of the $\mathbb{L}^2$ distance simplifies the problem, by forming an energy

$$E = \int_{[0,1]^2} \|\Phi(x) - (\phi_n \circ \cdots \circ \phi_1)(x)\|^2 dx \ . \tag{5}$$

We justify the use of $\mathbb{L}^2$ distance, over the geodesic distance, with the argument that minimization under one distance often leads to a minimizer under the other.

It is rather difficult to solve for all the parameters (for all the seeds) simultaneously. Indeed, the expression for cumulative deformation with just two local deformations gets complicated. Therefore, we take a sequential approach and add one local deformation to the model at a time. Let $\phi^{(k)} = \phi_k \circ \phi_{k-1} \circ \cdots \circ \phi_1$ be the cumulative deformation generated by first k seeds. Define two energy functions associated with this partial inference problem:

$$E^{(k+1)} = \int \left\| \Phi(x) - (\phi_{k+1} \circ \phi^{(k)})(x) \right\|^2 dx$$

and

$$\tilde{E}^{(k+1)} = \int \left\| \Phi(x) - \phi^{(k)}(x) - \psi_{k+1}(x) \right\|^2 dx \ .$$

$E^{(k+1)}$ denotes the energy under the composite model for $k+1$ seeds while $\tilde{E}^{(k+1)}$ denotes a similar energy except that the contribution from the last seed is considered additive. (Since this last seed is additive, it is relatively easier to solve for its parameters by minimizing $\tilde{E}^{(k+1)}$.) Our iterative approach is to solve for the parameters of ϕ_{k+1} to minimize $E^{(k+1)}$, for $k = 1, 2, \ldots, n$, and we will do so using a gradient approach. Similar to any gradient-based solution, the initialization of parameters becomes very important. For the purpose of improving initialization, we will solve for the parameters of ϕ_{k+1} under $\tilde{E}^{(k+1)}$ first and use these values as initial conditions in optimization of $E^{(k+1)}$.

We summarize the iterative procedure for estimating GRID parameters.

Algorithm 1 *Set $k = 0$.*

1. *Given the current estimated parameters for the k-seed composite model, $\{\theta_j\}_k = \{(\xi_j, a_j, c_j, \kappa_j, \tau_{0,j}) \mid j = 1, \ldots, k\}$, compute the cumulative deformation $\phi^{(k)}$.*
2. *Find $\theta_{k+1} = (\xi_{k+1}, a_{k+1}, c_{k+1}, \kappa_{k+1}, \tau_{0,k+1})$ for the $(k+1)^{th}$ seed by minimizing $\tilde{E}^{(k+1)}$. Use these values as initial condition for parameters of the $(k+1)^{th}$ seed.*
3. *For each possible permutation group of the set $\{1, 2, \ldots, k+1\}$, perform the following. Update each set of parameters $\{\theta_j\}_{k+1}$ using the gradient method to minimize $E^{(k+1)}$. Finally, choose the permutation/parameters that result in the minimum $E^{(k+1)}$.*
4. *Test the significance of the $(k+1)^{th}$ seed. If it is found significant, set $k = k+1$ and go to **step 1**; if not, **stop**.*

Note that even though the complexity of the composed deformation $\phi^{(k)}$ and thus $E^{(k)}$ increases with the number of seeds (or local elementary deformations), we still have analytical expressions for the gradients using the chain rule. It is important to note that all the previous seeds are re-estimated/updated as new seeds are added to the deformation. Thus, although this process is iterative, it is not incremental.

We have studied two cases for estimating ADFs:

1. **Constant ADF**: In the first case, we simplify the discussion by first assuming that $A(\tau) = a$ for all $\tau \in \mathbb{S}^1$. In this model, the partial derivatives of $E^{(k)}$ with respect to different parameters of ϕ_j for $j = 1, \cdots, k$ are given by:

$$\frac{\partial E^{(k)}}{\partial \xi_j^r} = \sum_{i=1}^{n} \frac{\partial E^{(k)}}{\partial \phi^{(k)}} \cdot \frac{\partial \phi^{(k)}}{\partial \phi^{(k-1)}} \cdot \frac{\partial \phi^{(k-1)}}{\partial \phi^{(k-2)}} \cdot \cdots \cdot \frac{\partial \phi^{(j+1)}}{\partial \phi^{(j)}} \cdot \frac{\partial \phi^{(j)}}{\partial \xi_j^r}, \quad r = 1, 2 \ ,$$

$$\frac{\partial E^{(k)}}{\partial a_j} = \sum_{i=1}^{n} \frac{\partial E^{(k)}}{\partial \phi^{(k)}} \cdot \frac{\partial \phi^{(k)}}{\partial \phi^{(k-1)}} \cdot \frac{\partial \phi^{(k-1)}}{\partial \phi^{(k-2)}} \cdot \cdots \cdot \frac{\partial \phi^{(j+1)}}{\partial \phi^{(j)}} \cdot \frac{\partial \phi^{(j)}}{\partial a_j} \ ,$$

$$\frac{\partial E^{(k)}}{\partial c_j} = \sum_{i=1}^{n} \frac{\partial E^{(k)}}{\partial \phi^{(k)}} \cdot \frac{\partial \phi^{(k)}}{\partial \phi^{(k-1)}} \cdot \frac{\partial \phi^{(k-1)}}{\partial \phi^{(k-2)}} \cdot \cdots \cdot \frac{\partial \phi^{(j+1)}}{\partial \phi^{(j)}} \cdot \frac{\partial \phi^{(j)}}{\partial c_j} \ .$$

   The different terms needed in these expressions, including $\frac{\partial E^{(j)}}{\partial \phi^{(j)}}$, $\frac{\partial \phi^{(j)}}{\partial \phi^{(j-1)}}$, $\frac{\partial \phi^{(j)}}{\partial \xi_j}$, $\frac{\partial \phi^{(j)}}{\partial a_j}$, and $\frac{\partial \phi^{(j)}}{\partial c_j}$, for $j = 1, \ldots, k$, are given in the appendix. Similar expressions can also be derived for the gradient of $\tilde{E}^{(k)}$ wrt the GRID parameters and those expressions are, as expected, simpler compared to the gradient given above.
2. **Non-isotropic ADF**: In the general case where the ADF is non-isotropic, we represent it using a scaled von-Mises density (the second term in Eqn. 3) and is parameterized by (a, κ, τ_0). For every deformation ϕ_j, the two additional parameters κ_j and $\tau_{0,j}$ are initialized in two steps. First, the ADF is estimated non-parametrically by integrating the deformation along each direction, as was done in [8]. Then, the parameters κ and τ_0 are estimated

from the nonparametric estimate using moment matching. The estimation is similar to estimating the parameters of von-Mises density from a sample by treating the non-parametric estimation as a weighted sample[17]. After these parameters are initialized, they are estimated using the gradient method, similar to the other parameters, with the gradient expression given in the appendix.

The last remaining item in Algorithm 1 is the test of significance of an incremental local deformation. In order to test the significance of the model with one more seed, several general methods for model selection may apply. Possible methods includes partial F test, AIC or BIC, and adjusted R^2. In this paper, a model is selected based on the adjusted R^2. It is a modification of R^2, which denotes the coefficient of determinant, that adjusts for the number of model parameters. Given any two estimated models, the model with the larger value of adjusted R^2 is preferred. In the experiments, we add one more seed if the improvement of adjusted R^2 is larger than a small cutoff value.

Since this method is based on a gradient search, it is difficult to claim a global solution. In principle, the solution obtained in the parameter space is a local one. However, there are some advantages to using this approach. Firstly, since the gradients of the energy function are available analytically, the gradient iterations are computationally fast. Secondly, for relatively small number of seeds in the model, the search over different orderings is efficient and gets us out of several local solutions.

4 Experimental Illustrations

Here we demonstrate the use of our framework for decomposing large cumulative deformations into smaller, parametric components using Algorithm 1. We will use both the simulated and real data to illustrate the estimation process, and will compare our results with those obtained using the additive model.

4.1 Synthetic Data

In order to validate estimation method for the composite model, we perform two experiments on the following types of synthetic data: (1) a 2D deformation with constant ADF, and (2) 2D deformation with non-constant ADF. In these experiments, a cumulative deformation is simulated using Eqn. 1 for an arbitrary number of local deformations, each with arbitrary parameters, and a white Gaussian noise is added at the end to form the observed deformation Φ.

Example 1: Fig. 2 shows the estimation results for the isotropic model. It shows the true underlying deformation (a), made up of $n = 2$ seeds, its noisy observation Φ (b) and several different GRID estimates from the observed Φ (c-e). Firstly, we estimate GRID components as described in Algorithm 1, with the result shown as $\hat{\Phi}_c$ in (e). Then, we reverse the order of two estimated seeds and

try to optimize their parameters, with the result shown as $\hat{\Phi'}_c$ in (c). The result of estimated deformation under the additive model is shown as $\hat{\Phi}_a$ in (d). To highlight the differences between different estimated deformations we also show their differences in the remaining panels. The Table 1 provides a quantification of estimation performance. It compares the energy E and estimated parameters for the three models, with the true values. Since the energy for $\hat{\Phi}_c$ is same as that for true underlying parameters, it shows the superiority of that estimation process described in Algorithm 1.

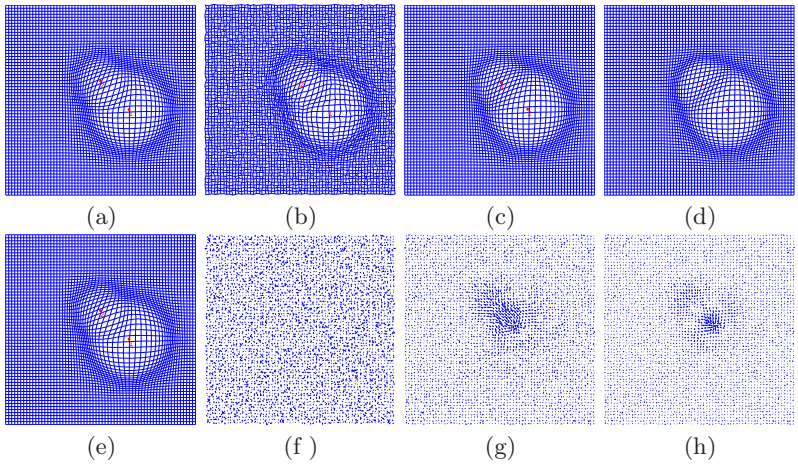


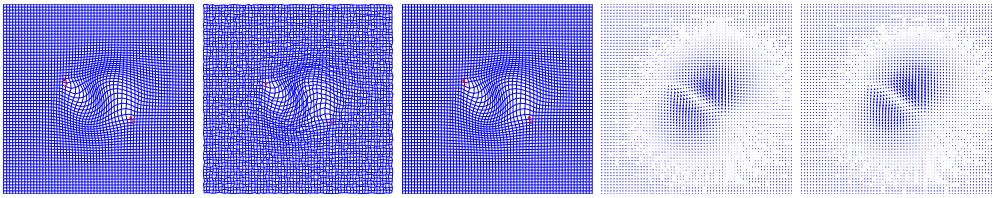
Fig. 2: Experiment with isotropic ADF: (a) synthetic deformation; (b)noisy observation Φ, and different estimated deformations (c) $\hat{\Phi}'_c$; (d)$\hat{\Phi}_a$, and (e)$\hat{\Phi}_c$. The differences (f) $\hat{\Phi}_c - \Phi$, (g)$\hat{\Phi}'_c - \Phi$, and (h)$\hat{\Phi}_a - \Phi$

Model	E	$\xi_1^{(1)}$	$\xi_1^{(2)}$	a_1	c_1	$\xi_2^{(1)}$	$\xi_2^{(2)}$	a_2	c_2
True	**0.0334**	0.6500	0.4500	1.0000	0.1500	0.5000	0.6000	0.8000	0.1000
Estimated $\hat{\Phi}_c$	**0.0334**	0.6500	0.4502	1.0045	0.1498	0.4995	0.6006	0.7947	0.0998
Reverse order $\hat{\Phi'}_c$	0.0576	0.5070	0.5930	0.7072	0.0956	0.6478	0.4525	1.0108	0.1527
Additive $\hat{\Phi}_a$	0.0548	0.6496	0.4507	1.0262	0.1501	0.5172	0.5827	0.7352	0.1004

Table 1: Estimation results for Example 1.

Example 2: Similarly, Fig. 3(a) shows an experiment involving seeds with non-isotropic ADFs. In this case we show the synthetic deformation (a) , its noisy version Φ (b), and the estimated deformation under the composite model $\hat{\Phi}_c$ (c). For further evaluation of this estimation, we show the true displacement

$\Psi(x) = \Phi(x) - x$ (d) and the estimated displacement $\hat{\Psi}$ (e).



(a) (b) (c) (d) (e)

Fig. 3: Experiment with non-isotropic ADF: (a)synthetic deformation; (b) noisy observation Φ; (c) estimated deformation map $\hat{\Phi}$; (d)true displacement field Ψ; (e)estimated displacement field $\hat{\Psi}$

Model	E	$\xi_1^{(1)}$	$\xi_1^{(2)}$	a_1	c_1	$\xi_2^{(1)}$	$\xi_2^{(2)}$	a_2	c_2
True	0.0330	0.6700	0.4000	1.0000	0.2000	0.3200	0.6000	1.0000	0.2000
Composite	0.0333	0.6677	0.4014	1.0004	0.1975	0.3233	0.5980	1.0248	0.1948
		κ_1	$\tau_{0,1}$			κ_2	$\tau_{0,2}$		
True		2.0000	2.3562			5.0000	5.4978		
Composite		1.9200	2.3379			4.9050	5.4865		

Table 2: Estimation Results for Example 2.

Ensemble Results: Beyond individual examples, we have exhaustively compared performances of the composite and the additive models using many realizations. Here we use 20 sets of data that are simulated from the 2D composite model with constant ADF and the parameters are then estimated using both the models. The results are summarized in Fig. 4. Panel (a) is a histogram of which model, composite or additive, is closer to the true deformation. Positive number indicates that the composite model outperforms the additive one, and vice-versa. We can see that for most data sets the composite model outperforms the additive model. Panel (b) presents a close up view of the parameter biases and variances. Each line represents the bias for one parameter. If the estimates equals the true parameters, the line will be all horizontal lines with $y = 0$. The top plot shows the biases from the additive model and the bottom one is for the composite model. The lines for composite model are all around zero with the ones for additive model having larger variation.

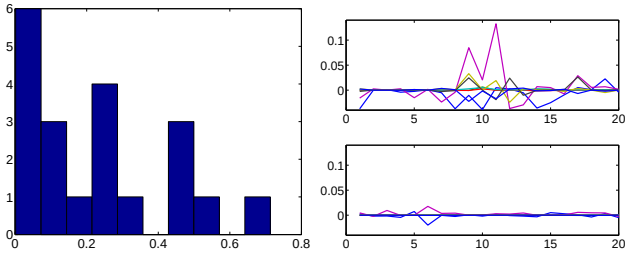


Fig. 4: (a) Relative errors under two models: $\frac{\|\hat{\Phi}_a - \Phi\| - \|\hat{\Phi}_c - \Phi\|}{\|\hat{\Phi}_a - \Phi\|}$; (b) Parameter estimation bias: top: $\hat{\theta}_a - \theta$, bottom: $\hat{\theta}_c - \theta$.

4.2 Estimation of GRID Component for Image Data

This method is applied to analysis of differences in human brains observed using MRI scans. In this case we perform the following experiment. We take two image scans as I and J, which are two slices of MRI of the same z coordinates from different subjects, and use a shape-based technique [10] to estimate a cumulative deformation Φ from I to J. This Φ is in the set of diffeomorphisms from $[0,1]^2$ to itself and minimizes a certain cost function involving $I \circ \Phi$ and J. Then, we use Algorithm 1, to estimate components of this deformation under the GRID model. We present two examples of this idea.

Example 1: Fig. 5 shows an example of images I and J and the deformation Φ that deforms I to match with J. In addition to plotting the maps $x \mapsto \Phi(x)$ as a surface mesh, one can also plot the displacement vector field $\Psi(x) = \Phi(x) - x$ for better visualization. Shown in the remaining panels are the estimated displacement vector $\hat{\Psi}$, showing the ordered sequence of displacement fields with the GRID components that were found using Algorithm 1, and a couple of ways of comparing the estimated deformations with the true deformations.
Example 2: Fig. 6 shows another example of estimating GRID parameters for deformation estimated from MRI images.

5 Conclusion

In this paper we have proposed a method to decompose and estimate the parameters in GRID based decomposition of anatomical deformations. The method preserves the iterative structure of the GRID model and gives an analytical form of the gradient for parameter estimation. Experimental results show that impacts from faraway seeds can be approximated by additive seeds model and composite model can not add much to it; however, as for seeds that are close to each other and have interaction, our method improves the estimation for large deformation as well as the model parameters.

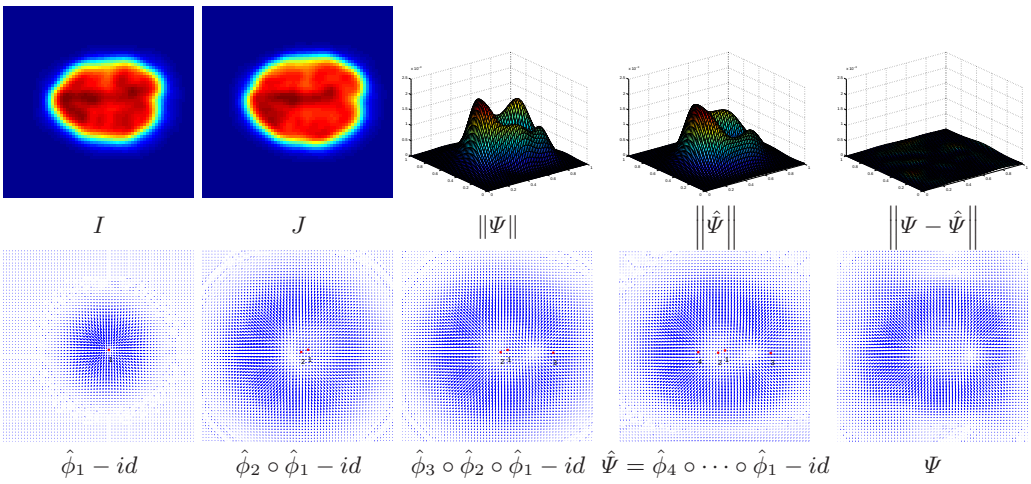


Fig. 5: Estimation of GRID components for deformation between MRI images.

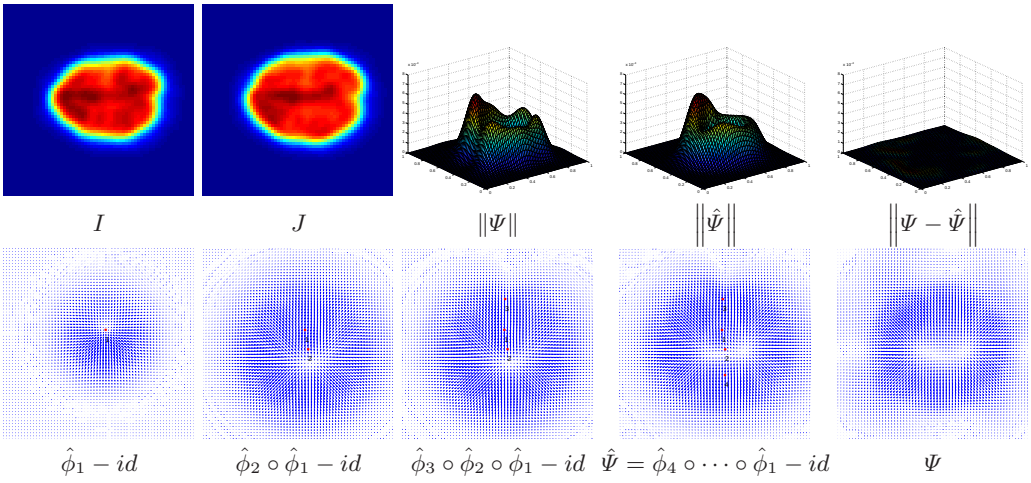


Fig. 6: Estimation of GRID components for deformation between MRI images.

References

1. Andresen, P.R., Nielsen, M.: Non-rigid registration by geometry-constraint diffusion. In: MICCAI. vol. 1679, pp. 533–543 (1999)
2. Arsigny, V., Pennec, X., Ayache, N.: Polyaffine transformations: a new class of diffeomorphisms for locally rigid or affine registration. In: Proc. Nat. Acad. Sci. vol. 90 (1993)

3. Chaplain, M.A.J.: Mathematical modeling of tumor growth. Springer, Seires: Interdisciplinary Applied Mathematics (2006)
4. Clatz, O., Bondiau, P.Y., Delingette, H., Malandain, G., Sermesant, M., Warfield, S.K., Ayache, N.: In silico tumor growth: application to glioblastomas. In: MICCAI. vol. 3217, pp. 337–345 (2004)
5. Durrleman, S., Prastawa, M., Gerig, G., Joshi, S.: Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. In: Inf. Process. Med. Imaging, vol. 22, pp. 123–134 (2011)
6. Grenander, U.: On the mathematics of growth. Quart. of Appl. Math. 65, 205–257 (2007)
7. Grenander, U., Miller, M.I.: Computational anatomy: an emerging discipline. Quart. Appl. Math LVI(4), 617–694 (1998)
8. Grenander, U., Srivastava, A., Saini, S.: A pattern-theoretic characterization of biological growth. IEEE Trans. Med. Imaging 26(5), 648–659 (2007)
9. Kansal, A.R., Torquato, S., IV, G.R.H., Chioccaeb, E.A., Deisboeck, T.S.: Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. J. Theor. Biol. 203, 367–382 (2000)
10. Kurtek, S., Klassen, E., Ding, Z., Jacobso, S., Jacobson, J., Avison, M., Srivastava, A.: Parameterization-invariant shape comparisons of anatomical surfaces. IEEE Trans. Med. Imaging 30(3), 849–858 (2011)
11. Ling, Y., He, B.: Entropic analysis of biological growth models. IEEE Trans. Biomed. Eng. 40(12), 1193–1200 (1993)
12. Miller, M., Christensen, G., Amit, Y., Grenander, U.: Mathematical textbook of deformable neuronatomies. Proc. Nat. Acad. Sci. 90(24), 11944–11948 (1993)
13. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: a general framework. Int. J. Comput. Vision 41(1/2), 61–84 (2001)
14. Portman, N., Grenander, U., Vrscay, E.R.: Direct estimation of biological growth properties from image data using the "grid" model. In: ICIAR '09. vol. 5627, pp. 832–843 (2009)
15. Portman, N., Grenander, U., Vrscay, E.R.: Grid macroscopic growth law and its application to image inference. Quart. Appl. Math. pp. 227–260 (2011)
16. Portman, N., Vrscay, E.R.: Existence and uniqueness of solutions to the grid macroscopic growth equation. Appl. Math. Comput. 217(21), 8318–8327 (2011)
17. Sra, S.: A short note on parameter approximation for Von Mises-Fisher distributions and a fast implementation of $i_s(x)$. Computational Stat. to appear (2009)
18. Srivastava, A., Saini, S., Ding, Z., Grenander, U.: Maximum-likelihood estimation of biological growth variables. In: EMMCVPR'05. vol. 3757, pp. 107–118 (2005)
19. Swanson, K.R., Bridge, C., Murray, J., Jr, E.C.A.: Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. J. Neurol. Sci. 216(1), 1–10 (2003)
20. Trouvé, A.: Diffeomorphisms groups and pattern matching in image analysis. Int. J. Comput. Vision 28(3), 213–221 (1998)

A Gradient Expressions

The terms, $\frac{\partial E^{(j)}}{\partial \phi^{(j)}} \in R^{1\times d}$, $\frac{\partial \phi^{(j)}}{\partial \phi^{(j-1)}} \in R^{d\times d}$, $\frac{\partial \phi^{(j)}}{\partial \xi_j} \in R^{d\times d}$, $\frac{\partial \phi^{(j)}}{\partial a_j}$, and $\frac{\partial \phi^{(j)}}{\partial c_j} \in R^{d\times 1}$, that are required in gradient computation are given as follows.

$$\frac{\partial E^{(j)}}{\partial \phi^{(j)}} = 2\left(\phi^{(j)} - \Phi\right) \ ,$$

$$\frac{\partial \phi^{(j)}}{\partial \phi^{(j-1)}} = a_{j-1} \cdot \exp\left(-\left(\phi^{(j-1)} - \xi_{j-1}\right)^2 / c_{j-1}^2\right) - 2a_{j-1} \cdot (\phi^{(j-1)} - \xi_{j-1})^2 / c_{j-1}^2 \cdot \exp\left(-\left(\phi^{(j-1)} - \xi_{j-1}\right)^2 / c_{j-1}^2\right) + 1 \ ,$$

$$\frac{\partial \phi^{(j)}}{\partial \xi_j} = a_j \cdot \exp\left(-\left(\phi^{(j-1)} - \xi_j\right)^2 / c_j^2\right) \cdot \left(2\left(\phi^{(j-1)} - \xi j\right)^2 / c_j^2 - 1\right) \ ,$$

$$\frac{\partial \phi^{(j)}}{\partial a_j} = (\phi^{(j-1)} - \xi_j) \cdot \exp\left(-\left(\phi^{(j-1)} - \xi_j\right)^2 / c_j^2\right) \ ,$$

$$\frac{\partial \phi^{(j)}}{\partial c_j} = 2a_j \cdot \exp\left(-\left(\phi^{(j-1)} - \xi_j\right)^2 / c_j^2\right) \cdot \left(\phi^{(j-1)} - \xi_j\right)^3 / c_j^3 \ .$$

When the ADF takes parametric forms of the second type, the gradient of the energy function with respect to the two parameters κ_j and $\tau_{0,j}$ for $j = 1, \cdots, k$ is calculated in the similar way as the other parameters. The partial derivatives from the one time deformation is shown as below.

$$\frac{\partial \phi^{(j)}}{\partial \kappa_j} = a_j \cdot \exp\left(\kappa_j \cos\left(\tau - \tau_{0,j}\right) - \kappa_j\right) \cdot \left(\phi^{(j-1)} - \xi_j\right) \cdot \exp\left(-\left(\phi^{(j-1)} - \xi_j\right)^2 / c_j^2\right) \cdot \left(\cos(\tau - \tau_{0,j}) - 1\right) \ ,$$

$$\frac{\partial \phi^{(j)}}{\partial \tau_{0,j}} = a_j \cdot \exp\left(\kappa_j \cos\left(\tau - \tau_{0,j}\right) - \kappa_j\right) \cdot \left(\phi^{(j-1)} - \xi_j\right) \cdot \exp\left(-\left(\phi^{(j-1)} - \xi_j\right)^2 / c_j^2\right) \cdot \kappa_j \cdot \sin\left(\tau - \tau_{0,j}\right) \ .$$

Teichmüller Shape Descriptor and its Application to Alzheimer's Disease Study

Wei Zeng[1], Rui Shi[1], Yalin Wang[2], and Xianfeng David Gu[1]

[1]Computer Science Department, SUNY at Stony Brook
[2]Computer Science and Engineering, Arizona State University

Abstract. We propose a novel method to apply Teichmüller space theory to study the signature of a family non-intersecting closed 3D curves on a general genus zero closed surface. Our algorithm provides an efficient method to encode both global surface and local contour shape information. The signature - Teichmüller shape descriptor - is computed by surface Ricci flow method, which is equivalent to solving an elliptic partial differential equation on surfaces and is quite stable. We propose to apply the new signature to analyze abnormalities in brain cortical morphometry. Experimental results with 3D MRI data from ADNI dataset (12 healthy controls versus 12 Alzheimer's disease (AD) subjects) demonstrate the effectiveness of our method and illustrate its potential as a novel surface-based cortical morphometry measurement in AD research.

1 Introduction

Some neurodegenerative diseases, such as Alzheimer's disease (AD), are characterized by progressive cognitive dysfunction. The underlying disease pathology most probably precedes the onset of cognitive symptoms by many years. Efforts are underway to find early diagnostic biomarkers to evaluate neurodegenerative risk presymptomatically in a sufficiently rapid and rigorous way. Among a number of different brain imaging, biological fluid and other biomarker measurements for use in the early detection and tracking of AD, structural magnetic resonance imaging (MRI) measurements of brain shrinkage are among the best established biomarkers of AD progression and pathology.

In structural MRI studies, early researches [30, 9] have demonstrated that surface-based brain mapping may offer advantages over volume-based brain mapping work [2] to study structural features of the brain, such as cortical gray matter thickness, complexity, and patterns of brain change over time due to disease or developmental processes. In research studies that analyze brain morphology, many surface-based shape analysis methods have been proposed, such as spherical harmonic analysis (SPHARM) [11, 4], minimum description length approaches [7], medial representations (M-reps) [24], cortical gyrification index [32], shape space [21], metamorphosis [33], momentum maps [25] and conformal invariants [34], etc.; these methods may be applied to analyze shape changes or abnormalities in cortical and subcortical brain structures. Even so, a stable method to compute a global intrinsic transformation-invariant shape descriptors would be highly advantageous in this research field.

Here, we propose a novel and intrinsic method to compute the global correlations between various surface region contours in Teichmüller space and apply it to study brain morphology in AD. The proposed shape signature demonstrates the global geometric features encoded in the interested regions, as a biomarker for measurements of AD progression and pathology. It is based on the brain surface conformal structure [18, 1, 13, 37] and can be accurately computed using the surface Ricci flow method [35, 20].

1.1 Related work

In brain mapping research, volumetric measures of structures identified on 3D MRI have been used to study group differences in brain structure and also to predict diagnosis [2]. Recent work has also used shape-based features [21, 33, 25], conformal invariants [34], analyzing surface changes using pointwise displacements of surface meshes, local deformation tensors, or surface expansion factors, such as the Jacobian determinant of a surface based mapping. For closed surfaces homotopic to a sphere, spherical harmonics have commonly been used for shape analysis, as have their generalizations, e.g., eigenfunctions of the Laplace-Beltrami operator in a system of spherical coordinates. These shape indices are also rotation invariant, i.e., their values do not depend on the orientation of the surface in space [30, 11, 28]]. Chung et al.[4] proposed a weighted spherical harmonic representation. For a specific choice of weights, the weighted SPHARM is shown to be the least squares approximation to the solution of an anisotropic heat diffusion on the unit sphere. Davies et al. performed a study of anatomical shape abnormalities in schizophrenia, using the minimal distance length approach to statistically align hippocampal parameterizations [7]. For classification, Linear Discriminant Analysis (LDA) or principal geodesic analysis can be used to find the discriminant vector in the feature space for distinguishing diseased subjects from controls. Tosun et al. [32] proposed the use of three different shape measures to quantify cortical gyrification and complexity. Gorczowski [12] presented a framework for discriminant analysis of populations of 3D multi-object sets. In addition to a sampled medial mesh representation, m-rep [24], they also considered pose differences as an additional statistical feature to improve the shape classification results.

For brain surface parameterization research, Schwartz et al. [26] and Timsari and Leahy [31] computed quasi-isometric flat maps of the cerebral cortex. Hurdal and Stephenson [18] reported a discrete mapping approach that uses circle packings to produce "flattened" images of cortical surfaces on the sphere, the Euclidean plane, and the hyperbolic plane. Angenent et al. [1] implemented a finite element approximation for parameterizing brain surfaces via conformal mappings. Gu et al. [13] proposed a method to find a unique conformal mapping between any two genus zero manifolds by minimizing the harmonic energy of the map. The holomorphic 1-form based conformal parameterization [37] can conformally parameterize high genus surfaces with boundaries but the resulting mappings have singularities. Other brain surface conformal parametrization methods, the Ricci flow method [35] and slit map method [36] can handle surfaces with complicated topologies (boundaries and landmarks) without singularities.

Wang et al. [34] applied the Yamabe flow method to study statistical group differences in a group of 40 healthy controls and 40 subjects with Williams syndrome, showing the potential of these surface-based descriptors for localizing cortical shape abnormalities in genetic disorders of brain development.

Conformal mappings have been applied in computer vision for modeling the 2D shape space by Sharon and Mumford [27]. The image plane is separated by a 2D contour, both interior and exterior are conformally mapped to disks, then the contour induces a diffeomorphism of the unit circle, which is the signature of the contour. The signature is invariant under translations and scalings, and able to recover the original contour by conformal welding. Later, this method is generalized to model multiple 2D contours with inner holes in [22]. To the best of our knowledge, our method is the first one to generalized Sharon and Mumford's 2D shape space to 3D surfaces.

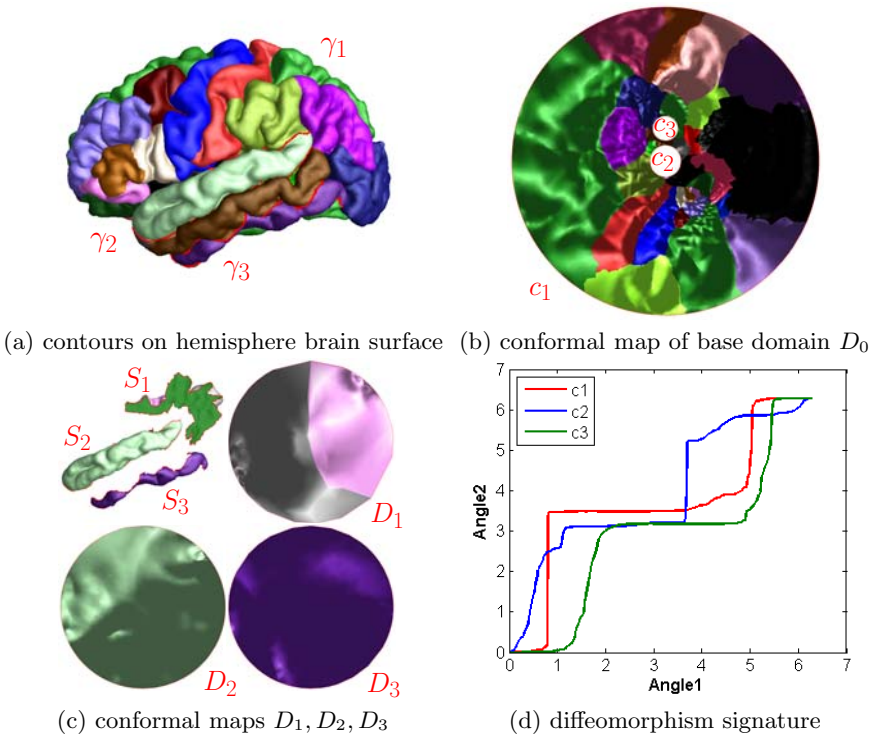
1.2 Our Approach

For a 3D surface, all the contours represent the 'shape' of the surface. Inspired by the beautiful research work of Sharon and Mumford [27] on 2D shape analysis (recently it has been generalized to model multiple 2D contours [22]), we build a Teichmüller space for 3D shapes by using conformal mappings. In this Teichmüller space, every 3D contour (a simple closed curve) is represented by a point in the space; each point denotes a unique equivalence class of diffeomorphisms up to a Möbius transformation. For a 3D surface, the diffeomorphisms of all the contours form a global shape representation of the surface. By using this signature, the similarities of 3D shapes can be quantitatively analyzed, therefore, the classification and recognition of 3D objects can be performed from their observed contours.

We tested our algorithm in some segmented regions on a set of brain left cortical surfaces extracted from 3D anatomical brain MRI scans. The proposed method can reliably compute signatures on two cortical functional areas by computing the diffeomorphisms of each observed contour. Using the signature as the statistics, our method achieve about 92% accuracy rate to discriminate a set of AD subjects from healthy control subjects.

To the best of our knowledge, it is the first work to apply contour diffeomorphism to brain morphometry research. Our experimental results demonstrated that this novel and simple method may be useful to analyze certain functional areas, and it may shed some lights on understanding detecting abnormality regions in brain surface morphometry. Our major **contributions** in this work include:

1. A new method to compute Teichmüller shape descriptor, in a way that generalized a prior 2D domain conformal mapping work [27].
2. The method is theoretically rigorous and general. It presents a stable way to calculate the diffeomorphisms of contours in general 3D surfaces based on Ricci flow.
3. It involves solving elliptic partial differential equations (PDEs), so it is numerically efficient and computationally stable.



(a) contours on hemisphere brain surface (b) conformal map of base domain D_0

(c) conformal maps D_1, D_2, D_3 (d) diffeomorphism signature

Fig. 1. Diffeomorphism signature via uniformization mapping for a genus zero surface with 3 simple closed contours $\gamma_1, \gamma_2, \gamma_3$ in (a), which correspond to the boundaries c_1, c_2, c_3 of the circle domains D_1, D_2, D_3 in (c), respectively. These three contours are also mapped to the boundaries of the base circle domain D_0 in (b). The curves in (d) demonstrate the diffeomorphisms for the three contours.

4. The shape descriptors are global and invariant to rigid motion and conformal deformations.

Pipeline. Figure 1 shows the pipeline for computing the diffeomorphism signature for a surface with 3 closed contours. Here, we use a human brain hemisphere surface whose functional areas are divided and labeled in different color. The contours (simple closed curves) of functional areas can be used to slice the surface open to connected patches. As shown in frames (a-c), three contours $\gamma_1, \gamma_2, \gamma_3$ are used to divide the whole brain (a genus zero surface S) to 4 patches S_0, S_1, S_2, S_3; each of them is conformally mapped to a circle domain (e.g., disk or annuli), D_0, D_1, D_2, D_3. Note that γ_1 is the contour of the joint functional areas of precuneus and posterior cingulate. One contour is mapped to two unit circles in two mappings. The representation of the shape according to each contour is a diffeomorphism of the unit circle to itself, defined as the mapping between periodic polar angles $(Angle_1, Angle_2), Angle_1, Angle_2 \in [0, 2\pi]$. The proper normalization is employed to remove Möbius ambiguity. The diffeomorphisms induced by the conformal maps of each curve form a diffeomorphism signature, which is the Teichmüller coordinates in Teichmüller space. As shown

in (d), the curves demonstrate the diffeomorphisms for three contours; the area distance is defined as the metric for shape comparison and classification.

2 Theoretical Background

In this section, we briefly introduce the theoretical foundations necessary for the current work. For more details, we refer readers to the classical books [10, 16].

2.1 Surface Uniformization Mapping

Conformal mapping between two surfaces preserves angles. Suppose $(S_1, \mathbf{g}_1)$ and $(S_2, \mathbf{g}_2)$ are two surfaces embedded in $\mathbb{R}^3$, $\mathbf{g}_1$ and $\mathbf{g}_2$ are the Euclidean induced Riemannian metrics. A mapping $\phi : S_1 \to S_2$ is called *conformal*, if the pull back metric of $\mathbf{g}_2$ induced by ϕ on S_1 differs from $\mathbf{g}_1$ by a positive scalar function: $\phi^* \mathbf{g}_2 = e^{2\lambda} \mathbf{g}_1$, where $\lambda : S_1 \to \mathbb{R}$ is a scalar function, called the *conformal factor*.

For example, all the conformal automorphisms of the unit disk form the *Möbius transformation group* of the disk, each mapping is given by

$$z \to e^{i\theta} \frac{z - z_0}{1 - \bar{z}_0 z}.$$

All the conformal automorphism group of the extended complex plane $\mathbb{C} \cup \{\infty\}$ is also called Möbius transformation group, each mapping is given by

$$z \to \frac{az + b}{cz + d}, ad - bc = 1, a, b, d, c \in \mathbb{C}.$$

By stereo-graphic projection, the unit sphere can be conformally mapped to the extended complex plane. Therefore, the Möbius transformation group is also the conformal automorphism group of the unit sphere.

A circle domain on the complex plane is the unit disk with circular holes. A circle domain can be conformally transformed to another circle domain by Möbius transformations, $z \to e^{i\theta} \frac{z - z_0}{1 - \bar{z}_0 z}$. All genus zero surfaces with boundaries can be conformally mapped to circle domains:

Theorem 21 (Uniformization) *Suppose S is a genus zero Riemannian surface with boundaries, then S can be conformally mapped onto a circle domain. All such conformal mappings differ by a Möbius transformation on the unit disk.*

This theorem can be proved using Ricci flow straightforwardly. Therefore, the conformal automorphism group of S $Conf(S)$ is given

$$Conf(S) := \{\phi^{-1} \circ \tau \circ \phi | \tau \in M\ddot{o}b(\mathbb{S}^2)\}.$$

2.2 Teichmüller Space

Definition 22 (Conformal Equivalence) *Suppose $(S_1, \mathbf{g}_1)$ and $(S_2, \mathbf{g}_2)$ are two Riemannian surfaces. We say S_1 and S_2 are* conformally equivalent *if there is a conformal diffeomorphism between them.*

All Riemannian surfaces can be classified by the conformal equivalence relation. Each conformal equivalence class shares the same *conformal invariants*, the so-called *conformal module*. The conformal module is one of the key component for us to define the unique shape signature.

Definition 23 (Teichmüller Space) *Fixing the topology of the surfaces, all the conformal equivalence classes form a manifold, which is called the Teichmüller space.*

For example, all topological disks (genus zero Riemannian surfaces with single boundary) can be conformally mapped to the planar disk. Therefore, the Teichmuller space for topological disks consists of a single point.

Suppose a genus zero Riemannian surface S has n boundary components $\{\gamma_1, \gamma_2, \cdots, \gamma_n\}$, $\partial S = \gamma_1 + \gamma_2 + \cdots + \gamma_n$, $\phi : S \to \mathbb{D}$ is the conformal mapping that maps S to a circle domain $\mathbb{D}$, such that (a). $\phi(\gamma_1)$ is the exterior boundary of the $\mathbb{D}$; (b) $\phi(\gamma_2)$ centers at the origin; (c) The center of $\phi(\gamma_3)$ is on the imaginary axis. Then the conformal module of the surface S (also the circle domain $\mathbb{D}$) is given by $Mod(S) = \{(c_i, r_i)|i = 1, 2, \cdots, n\}$. This shows the Teichmüller space of genus zero surfaces with n boundaries is of $3n - 6$ dimensional. The Teichmüller space has a so-called Weil-Peterson metric [27], so it is a Riemannian manifold. Furthermore it is with negative sectional curvature, therefore, the geodesic between arbitrary two points is unique.

2.3 Surface Ricci Flow

Surface Ricci flow is the powerful tool to compute uniformization. *Ricci flow* refers to the process of deforming Riemannian metric $\mathbf{g}$ proportional to the curvature, such that the curvature K evolves according to a heat diffusion process, eventually the curvature becomes constant everywhere. Suppose the metric $\mathbf{g} = (g_{ij})$ in local coordinate. Hamilton [15] introduced the Ricci flow as

$$\frac{dg_{ij}}{dt} = -Kg_{ij}.$$

Surface Ricci flow conformally deforms the Riemannian metric, and converges to constant curvature metric [3]. Furthermore, Ricci flow can be used to compute the unique conformal Riemannian metric with the prescribed curvature.

Theorem 24 (Hamilton and Chow [3]) *Suppose S is a closed surface with a Riemannian metric. If the total area is preserved, the surface Ricci flow will converge to a Riemannian metric of constant Gaussian curvature.*

2.4 Teichmüller Shape Descriptor

Suppose $\Gamma = \{\gamma_0, \gamma_1, \cdots, \gamma_n\}$ is a set of non-intersecting smooth closed curves on a genus zero closed surface. Γ segments the surface to a set of connected components $\{\Omega_0, \Omega_1, \cdots, \Omega_n\}$, each segment Ω_i is a genus zero surface with boundary components. Construct the uniformization mapping $\phi_k : \Omega_k \to \mathbb{D}_k$

to map each segment Ω_k to a circle domain $\mathbb{D}_k$, $0 \leq k \leq n$. Assume γ_i is the common boundary between Ω_j and Ω_k, then $\phi_j(\gamma_i)$ is a circular boundary on the circle domain $\mathbb{D}_j$, $\phi_k(\gamma_i)$ is another circle on $\mathbb{D}_k$. Let $f_i|_{\mathbb{S}^1} := \phi_j \circ \phi_k^{-1}|_{\mathbb{S}^1} : \mathbb{S}^1 \to \mathbb{S}^1$ be the diffeomorphism from the circle to itself. We called the the diffeomorphism f_i the *signature of* γ_i.

Definition 25 (Signature of a Family of Loops) *The signature of a family non-intersecting closed 3D curves* $\Gamma = \{\gamma_0, \gamma_1, \cdots, \gamma_k\}$ *on a genus zero closed surface is defined as:* $S(\Gamma) := \{f_0, f_1, \cdots, f_k\} \cup \{Mod(\mathbb{D}_0), Mod(\mathbb{D}_1), \cdots, Mod(\mathbb{D}_k)\}$.

The following **main theorem** plays fundamental role for the current work. Note that if a circle domain $\mathbb{D}_k$ is disk, its conformal module can be omitted from the signature.

Theorem 26 (Main Theorem) *The family of smooth 3D closed curves* Γ *on a genus zero closed Riemannian surface is determined by its signature* $S(\Gamma)$*, unique up to a conformal automorphism of the surface* $\eta \in Conf(S)$.

The proof of Theorem 26 can be found in the appendix section.

The theorem states that the proposed signature determine shapes up to a Möbius transformation. We can further do a normalization that fixes ∞ to ∞ and that the differential carries the real positive axis at ∞ to the real positive axis at ∞, as in Sharon and Mumford's paper [27]. The signature can then determine the shapes uniquely up to translation and scaling.

The shape signature $S(\Gamma)$ gives us a complete representation for the space of shapes. It inherits a natural metric. Given two shapes Γ_1 and Γ_2. Let $S(\Gamma_i) := \{f_0^i, f_1^i, \cdots, f_k^i\} \cup \{Mod(\mathbb{D}_0^i), Mod(\mathbb{D}_1^i), \cdots, Mod(\mathbb{D}_k^i)\}$ $(i = 1, 2)$. We can define a metric $d(S(\Gamma_1), S(\Gamma_2))$ between the two shape signatures using the natural metric in the Teichmüller space.Our signature is stable under geometric noise. Our algorithm depends on conformal maps from surfaces to circle domains using discrete Ricci flow method.

3 Algorithm

In this section, we explain each step of the pipeline in Figure 1 in details.

3.1 Circular Uniformization Mapping

We apply discrete Ricci flow method [20] to conformally map the surfaces onto planar circle domains $\phi_k : S_k \to \mathbb{D}$. The surface is represented as a triangle mesh Σ. A discrete Riemannian metric is represented as the edge length.

We associate each vertex v_i with a circle (v_i, γ_i), where γ_i is the radius. Let $u_i = \log \gamma_i$ be the discrete conformal factor. The discrete Ricci flow is defined as follows:

$$\frac{du_i(t)}{dt} = (\bar{K}_i - K_i), \tag{1}$$

where $\bar{K}_i$ is the user defined target curvature and K_i is the curvature induced by the current metric. The discrete Ricci flow has exactly the same form as the

smooth Ricci flow, which conformally deforms the discrete metric according to the Gaussian curvature. The computation is based on circle packing metric [20].

Suppose Σ is a genus zero mesh with multiple boundary components. The uniformization conformal mapping $\phi : \Sigma \to \mathbb{D}$, where $\mathbb{D}$ is the circle domain, can be computed using Ricci flow by setting the prescribed curvature as follows: (a) The geodesic curvature on the exterior boundary is $+1$ everywhere; (b) the geodesic curvature on other boundaries are negative constants; (c) the Gaussian curvature on interior points are zeros everywhere. We use this method to compute conformal mapping, and get conformal module and shape descriptor. The main challenge is that the target curvature is dynamically determined by the metric. The metric is evolving, so is the target curvature. The detailed algorithm is reported in [38].

3.2 Computing Shape Descriptor

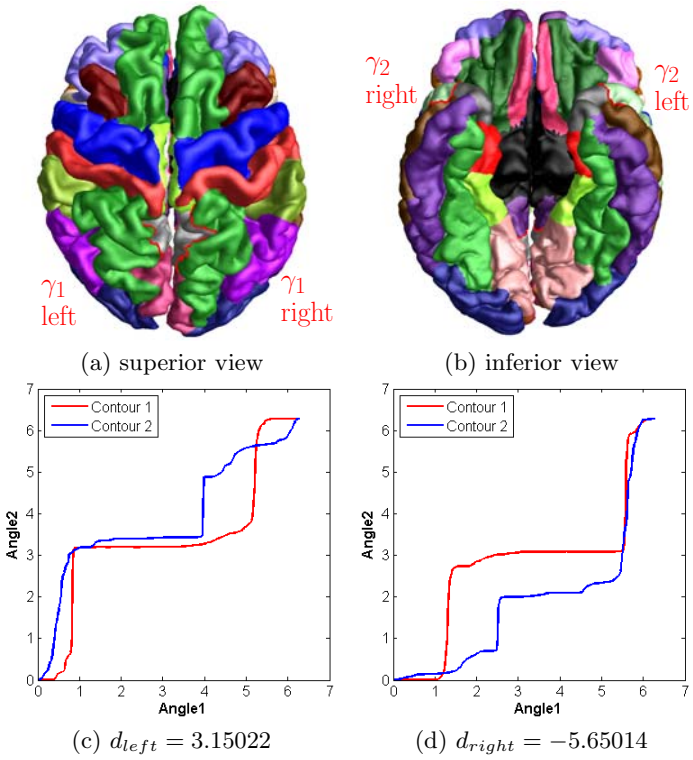
After the computation of the conformal mapping, each connected component is mapped to a circle domain. We define an order for all the loops on the surface, this induces an order for all the boundary components on each segment. Then by the definition for the conformal module of a circle domain, we normalize each circle domain using a Möbius transformation, then compute the conformal modules directly. For those segments, which are simply connected and mapped to the unit disk, we compute its mass center, and use a Möbius transformation to map the center to the origin.

Each loop on the surface becomes the boundary components on two segments, both boundary components are mapped to a circle under the uniformization mapping. Then we compute the signature directly.

4 Experimental Results

We demonstrate the efficiency and efficacy of our method by analyzing the human brain cortexes of Alzheimer's disease (AD) and healthy control subjects. The brain surfaces are represented as triangular meshes; a half brain with $100K$ triangles. We implement the algorithm using generic C++ on windows XP platform, with Intel Xeon CPU 3.39 GHz, 3.98 G RAM. The numerical systems are solved using Matlab C++ library. In general, the signature calculation on each half brain surface with 2 or 3 contours on each half takes less than 1 minute to compute, even on complicated domains.

Data and preprocessing. The experimental data include 12 Alzheimer disease patients and 12 healthy control subjects. The structural MRI images were from the AD Neuroimaging Initiative (ADNI [19, 23]. We used Freesurfer's automated processing pipeline [6] for automatic skull stripping, tissue classification, surface extraction, cortical and subcortical parcellations. It calculates volumes of individual grey matter parcellations in mm^3 and surface area in mm^2. It also provides surface and volume statistics for about 34 different cortical structures, and also computes geometric characteristics such as curvature, curvedness, local foldedness for each of the parcellations [8]. In this work, we studied segmented surface regions for group difference analysis.



(a) superior view (b) inferior view

(c) $d_{left} = 3.15022$ (d) $d_{right} = -5.65014$

Fig. 2. Diffeomorphisim signature (d_{left}, d_{right}) of a healthy control brain cortex. Each ($left$ and $right$) half brain is a genus zero surface with 2 contours.

Quantitative analysis. Figure 2 shows an example of diffeomorphism signatures for a brain cortical surface. We selected two contours on the left and the right half brain cortical surfaces, which correspond to superior temporal and the joint areas of precuneus and posterior cingulate. Early researches [17, 14] have indicated that these two areas may have significant atrophy in AD group. These two contours segment a brain hemisphere surface to 3 patches; one topological annulus (called the base domain), two topological disks. The base domain with two boundaries is mapped to an annulus, one boundary to exterior unit circle, the other one to the inner concentric circle. The diffeomorphism signature for each contour is plotted as a monotonic curve within the square $[0, 2\pi] \times [0, 2\pi]$. The area difference between the plotted curves, $d = \int_0^{2\pi} (Angle_2^2 - Angle_2^1) dAngle_1$, is used as the metric to represent the global shape of both contours. So the signature of the whole brain surface is represented as a pair (d_{left}, d_{right}) for combining the left-hemisphere and right-hemisphere brain shape signatures. The method was tested on 12 AD subjects and 12 healthy subjects, with mean signatures $(3.6827, -7.12957)$ and $(5.2752, -5.6036)$, respectively. Figure 3 shows that with a simple linear discriminant analysis (LDA) model, there were only two subjects that were not correctly classified. It demonstrates that the proposed global diffeomorphism signature of contours is very efficient and may be effective to differentiate the shapes within healthy control and AD subject groups.

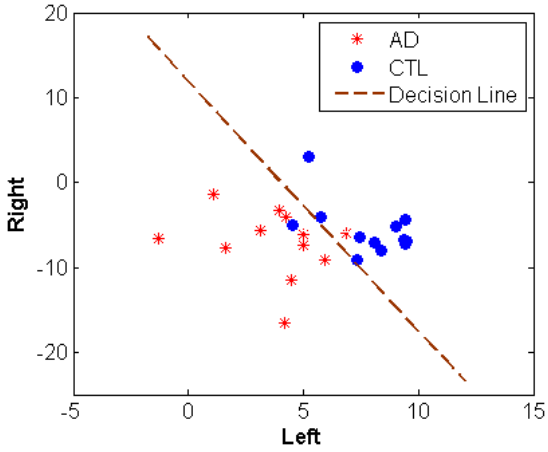


Fig. 3. Distribution of diffeomorphism signature for 12 AD (in red) and 12 healthy control (CTL) (in blue) subjects. Each point denotes the diffeomorphism signature value (d_{left}, d_{right}) for a whole brain surface, computed as in Figure 2.

Discussion. The proposed work is based on surface Ricci flow research. Computing the conformal module is equivalent to solving an elliptic partial differential equation on surfaces. According to PDE theory, the solution is smoother than its boundary conditions, so the solution process is quite stable.

For surface-based AD research, the state-of-the-art work has used cortical thickness as the measurement [29, 5]. However, recent research [39] indicated that the commonly used cortical thickness and cortical area measurements are genetically and phenotypically independent. The biological meaning of the proposed shape signature is closely related to brain atrophy so it is more related to cortical area changes. Our method provides a unique and intrinsic shape feature to study brain morphometry changes caused by brain atrophy. It studies the sensitivity and reproducibility of shape features computed in the entire brain surface domain. The gained insights help improve our understanding to AD related pathology and discover the precise etiology of the grey matter changes. The preliminary results demonstrated that the shape signature provides a reasonably good discriminant power for AD biomarker research. We currently studied the superior temporal area, which is directly related to medial temporal lobe atrophy. The method can be equally applied to other regions as well. In future, we may study/compare other functional areas in the medial temporal lobe.

5 Conclusion

In this paper, we propose a novel method that computes the global shape signatures on specified functional areas on brain cortical surfaces in Teichmüller space. In the future, we will further explore and validate other applications of this global correlation shape signature in neuroimaging and shape analysis research.

References

1. S. Angenent, S. Haker, R. Kikinis, and A. Tannenbaum. Nondistorting flattening maps and the 3D visualization of colon CT images. *IEEE Trans. Med. Imag.*, 19:665–671, 2000.
2. J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston. Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping*, 6:348–357, 1998.
3. B. Chow, P. Lu, and L. Ni. *Hamilton's Ricci Flow*. American Mathematical Society, 2006.
4. M. K. Chung, K. M. Dalton, and R. J. Davidson. Tensor-based cortical surface morphometry via weighted spherical harmonic representation. *IEEE Trans. Med. Imag.*, 27:1143–1151, 2008.
5. R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766 – 781, 2011.
6. A. M. Dale, B. Fischl, and M. I. Sereno. Cortical surface-based analysis I: segmentation and surface reconstruction. *Neuroimage*, 27:179–194, 1999.
7. R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor. Shape discrimination in the hippocampus using an MDL model. *in Proc. Infor. Proc. Med. Imag. (IPMI)*, 2003.
8. R. S. Desikan, F. Segonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31:968–80, 2006.
9. B. Fischl, M. I. Sereno, and A. M. Dale. Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9:195 – 207, 1999.
10. F. P. Gardiner and N. Lakic. *Quasiconformal Teichmuller theory*. American Mathematical Society, 2000.
11. G. Gerig, M. Styner, D. Jones, D. Weinberger, and J. Lieberman. Shape analysis of brain ventricles using SPHARM. *in Proc. MMBIA 2001*, pages 171–178, 2001.
12. K. Gorczowski, M. Styner, J.-Y. Jeong, J. S. Marron, J. Piven, H. C. Hazlett, S. M. Pizer, and G. Gerig. Statistical shape analysis of multi-object complexes. *IEEE Conf. Comp. Vis. Patt. Recog. CVPR '07*, pages 1–8, 2007.
13. X. Gu, Y. Wang, T. F. Chan, P. M. Thompson, and S.-T. Yau. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Trans. Med. Imag.*, 23:949–958, 2004.
14. X. Guo, Z. Wang, K. Li, Z. Li, Z. Qi, Z. Jin, L. Yao, and K. Chen. Voxel-based assessment of gray and white matter volumes in Alzheimer's disease. *Neurosci Lett*, 468:146–50, 2010.
15. R. S. Hamilton. The Ricci flow on surfaces. *Mathematics and general relativity*, 71:237–262, 1988.
16. P. Henrici. *Applied and Computational Complex Analysis*, volume 3. Wiley-Intersecience, 1988.
17. X. Hua, S. Lee, D. P. Hibar, I. Yanovsky, A. D. Leow, A. W. Toga, C. R. J. Jr, M. A. Bernstein, E. M. Reiman, D. J. Harvey, J. Kornak, N. Schuff, G. E. Alexander, M. W. Weiner, and P. M. Thompson. Mapping Alzheimer's disease progression in 1309 MRI scans: Power estimates for different inter-scan intervals. *Neuroimage*, 51:63–75, 2010.
18. M. K. Hurdal and K. Stephenson. Cortical cartography using the discrete conformal approach of circle packings. *NeuroImage*, 23:S119–S128, 2004.

19. C. R. J. Jack, M. A. Bernstein, N. C. Fox, P. M. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, and e. al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. of Mag. Res. Ima.*, 27:685–691, 2007.
20. M. Jin, J. Kim, F. Luo, and X. Gu. Discrete surface Ricci flow. *IEEE Trans. Vis. Comput. Graphics*, 14:1030–1043, September 2008.
21. X. Liu, Y. Shi, I. Dinov, and W. Mio. A computational model of multidimensional shape. *Int J Comput Vis*, 89:69–83, 2010.
22. L. M. Lui, W. Zeng, S.-T. Yau, and X. Gu. Shape analysis of planar objects with arbitrary topologies using conformal geometry. In *ECCV 2010*, 2010.
23. S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett. The Alzheimer's disease neuroimaging initiative. *Neuroimaging clinics of North America*, 15:869–877, 2005.
24. S. Pizer, D. Fritsch, P. Yushkevich, V. Johnson, and E. Chaney. Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Trans. Med. Imag.*, 18:851–865, 1999.
25. A. Qiu and M. I. Miller. Multi-structure network shape analysis via normal surface momentum maps. *NeuroImage*, 42:1430–8, 2008.
26. E. L. Schwartz, A. Shaw, and E. Wolfson. A numerical solution to the generalized Mapmaker's problem: Flattening nonconvex polyhedral surfaces. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11:1005–1008, 1989.
27. E. Sharon and D. Mumford. 2D-shape analysis using conformal mapping. *Int. J. Comput. Vision*, 70:55–75, October 2006.
28. L. Shen, A. J. Saykin, M. K. Chung, and H. Huang. Morphometric analysis of hippocampal shape in mild cognitive impairment: An imaging genetics study. *in IEEE 7th International Conference Bioinformatics and Bioengineering*, 2007.
29. P. M. Thompson, K. M. Hayashi, G. D. Zubicaray, A. L. Janke, S. E. Rose, J. Semple, D. Herman, M. S. Hong, S. S. Dittmer, D. M. Doddrell, and A. W. Toga. Dynamics of gray matter loss in Alzheimer's disease. *J. Neuroscience*, 23:994–1005, 2003.
30. P. M. Thompson and A. W. Toga. A surface-based technique for warping 3-dimensional images of the brain. *IEEE Trans. Med. Imag.*, 15:1–16, 1996.
31. B. Timsari and R. M. Leahy. Optimization method for creating semi-isometric flat maps of the cerebral cortex. *Medical Imaging 2000: Image Processing*, 3979:698–708, 2000.
32. D. Tosun, A. Reiss, A. D. Lee, R. A. Dutton, K. M. Hayashi, U. Bellugi, A. M. Galaburda, J. R. Korenberg, D. L. Mills, A. W. Toga, and P. M. Thompson. Use of 3-D cortical morphometry for mapping increased cortical gyrification and complexity in Williams syndrome. *Biom. Imag.: Nano to Macro, 2006. 3rd IEEE Int. Symp. on, 2006*, pages 1172–1175, 2006.
33. A. Trouve and L. Younes. Metamorphoses through Lie group action. *Foundations of Computational Mathematics*, 5:173–198, 2005.
34. Y. Wang, X. Gu, T. F. Chan, and P. M. Thompson. Shape analysis with conformal invariants for multiply connected domains and its application to analyzing brain morphology. *IEEE Conf. Comp. Vis. Patt. Recog. CVPR '09*, pages 202–209, 2009.
35. Y. Wang, X. Gu, T. F. Chan, P. M. Thompson, and S.-T. Yau. Brain surface conformal parameterization with algebraic functions. *Med. Image Comp. Comput.-Assist. Intervention, Proceedings, Part II*, pages 946–954, 2006.
36. Y. Wang, X. Gu, T. F. Chan, P. M. Thompson, and S.-T. Yau. Conformal slit mapping and its applications to brain surface parameterization. *Med. Image Comp. Comput.-Assist. Intervention, Proceedings, Part I*, pages 585–593, 2008.
37. Y. Wang, L. Lui, X. Gu, K. M. Hayashi, T. F. Chan, A. W. Toga, P. M. Thompson, and S.-T. Yau. Brain surface conformal parameterization using Riemann surface structure. *IEEE Trans. Med. Imag.*, 26:853–865, 2007.

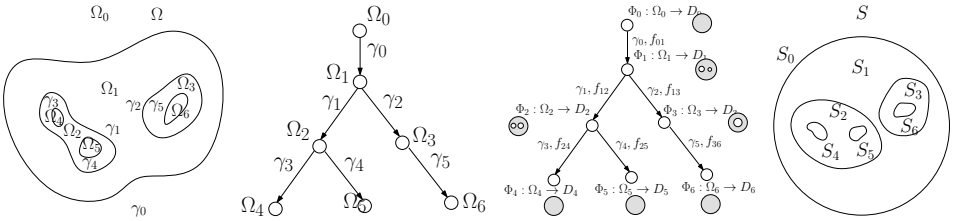


Fig. 4. Proof for the main theorem, the signature uniquely determines the family of closed curves unique up to a Mb̈oius transformation.

38. Y. Wang, J. Shi, X. Yin, X. Gu, T. F. Chan, S.-T. Yau, A. W. Toga, and P. M. Thompson. Brain surface conformal parameterization with the Ricci flow. *submitted to IEEE Trans. Med. Imag.*, 2011.
39. A. M. Winkler, P. Kochunov, J. Blangero, L. Almasy, K. Zilles, P. T. Fox, R. Duggirala, and D. C. Glahn. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, 53(3):1135 – 1146, 2010.

Appendix: Proof of Theorem 26

proof See Figure 4. In the left frame, a family of planar smooth curves $\Gamma = \{\gamma_0, \cdots, \gamma_5\}$ divide the plane to segments $\{\Omega_0, \Omega_1, \cdots, \Omega_6\}$, where Ω_0 contains the ∞ point. We represent the segments and the curves as a tree in the second frame, where each node represents a segment Ω_k, each link represents a curve γ_i. If Ω_j is included by Ω_i, and Ω_i and Ω_j shares a curve γ_k, then the link γ_k in the tree connects Ω_j to Ω_i, denoted as $\gamma_k : \Omega_i \to \Omega_j$. In the third frame, each segment Ω_k is mapped conformally to a circle domain D_k by Φ_k. The signature for each closed curve γ_k is computed $f_{ij} = \Phi_i \circ \Phi_j^{-1}|_{\gamma_k}$, where $\gamma_k : \Omega_i \to \Omega_j$ in the tree. In the last frame, we construct a Riemann sphere by gluing circle domains D_k's using f_{ij}'s in the following way. The gluing process is of bottom up. We first glue the leaf nodes to their fathers. Let $\gamma_k : D_i \to D_j$, D_j be a leaf of the tree. For each point $z = re^{i\theta}$ in D_j, the *extension map*: $G_{ij}(re^{i\theta}) = re^{f_{ij}(\theta)}$.

We denote the image of D_j under G_{ij} as S_j. Then we glue S_j with D_i. By repeating this gluing procedure bottom up, we glue all leafs to their fathers. Then we prune all leaves from the tree. Then we glue all the leaves of the new tree, and prune again. By repeating this procedure, eventually, we get a tree with only the root node, then we get a Riemann sphere, denoted as S. Each circle domain D_k is mapped to a segment S_k in the last frame, by a sequence of extension maps. Suppose D_k is a circle domain, a path from the root D_0 to D_k is $\{i_0 = 0, i_1, i_2, \cdots, i_n = k\}$, then the map from $G_k : D_k \to S_k$ is given by: $G_k = G_{i_0 i_1} \circ G_{i_1 i_2} \circ \cdots \circ G_{i_{n-1} i_n}$. Note that, G_0 is identity. Then the Beltrami coefficient of $G_k^{-1} : S_k \to D_k$ can be directly computed, denoted as $\mu_k : S_k \to \mathbb{C}$. The composition $\Phi_k \circ G_k^{-1} : S_k \to \Omega_k$ maps S_k to Ω_k, because Φ_k is conformal, therefore the Beltrami coefficient of $\Phi_k \circ G_k^{-1}$ equals to μ_k.

We want to find a map from the Riemann sphere S to the original Riemann sphere Ω, $\Phi : S \to \Omega$. The Beltrami-coefficient $\mu : S \to \mathbb{C}$ is the union of μ_k's

each segments: $\mu(z) = \mu_k(z), \forall z \in S_k$. The solution exists and is unique up to a Möbius transformation according to Quasi-conformal Mapping theorem [10].

Note that, the discrete computational method is more direct without explicitly solving the Beltrami equation. From the Beltrami coefficient μ, one can deform the conformal structure of S_k to that of Ω_k, under the conformal structures of Ω_k, $\Phi : S \to \Omega$ becomes a conformal mapping. The conformal structure of Ω_k is equivalent to that of D_k, therefore, one can use the conformal structure of D_k directly. In discrete case, the conformal structure is represented as the angle structure. Therefore in our algorithm, we copy the angle structures of D_k's to S, and compute the conformal map Φ directly.

On a Geodesic Equation for Planar Conformal Template Matching

Stephen Marsland[1], Robert McLachlan[2], Klas Modin[1], and Matthew Perlmutter[2]

[1] School of Engineering and Advanced Technology (SEAT)
Massey University, Palmerston North, New Zealand
[2] Institute of Fundamental Sciences (IFS)
Massey University, Palmerston North, New Zealand

Abstract. In this paper we consider planar conformal deformations, motivated by the warps that Wentworth Thompson used to deform images of one species into another. We study an equation for geodesic motion on the infinite dimensional Fréchet manifold $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ of conformal embeddings of the disk into the plane. We demonstrate that solutions may be represented as sheets, and use the sheet ansatz to derive a numerical discretization scheme. We also show that the equation admits totally geodesic solutions corresponding to scaling and translation, but not to affine transformations.

1 Introduction

The use of diffeomorphic transformations in both image registration and shape analysis is now common and utilised in many machine vision and image analysis tasks. One image or shape is brought into alignment with another by deforming the image until some similarity measure reaches a minimum. The deformation is computed as a geodesic with respect to some metric on the diffeomorphism group. For a general treatment and an overview of the subject see [1] and references therein.

The standard approach to the deformation method is to first perform an affine registration (principally to remove translation and rotation), and then to seek a diffeomorphic warp of the image. However, in what is arguably the most influential demonstration of the application of warping methods – D'Arcy Wentworth Thompson's seminal book 'On Growth and Form' [2] – Thompson transforms images of one species into another using relatively simple warps, so that the gross features of the two match. In a recent review of his work, biologist Arthur Wallace says:

This theory cries out for causal explanation, which is something the great man eschewed. [...] His transformations suggest coordinated rather than piecemeal changes to development in the course of evolution, an issue which almost completely disappeared from view in the era of the 'modern synthesis' of evolutionary theory, but which is of central importance again in the era of evo-devo.

[. . .] All the tools are now in place to examine the mechanistic basis of transformations. Not only do we have phylogenetic systematics and evo-devo, but, so obvious that it is easy to forget, we have computers, and especially, in this context, advanced computer graphics. We owe it to the great man to put these three things together to investigate the mechanisms that produce the morphological changes that he captured so elegantly with little more than sheets of graph paper and, of course, a brilliant mind. [3]

Figure no. in [2]	Transformation group
515	$x \mapsto ax,\ y \mapsto y$
513.2	$x \mapsto ax,\ y \mapsto by$
509, 510, 518	$x \mapsto ax,\ y \mapsto cx + dy$ (shears)
521–22, 513.5	$x \mapsto ax + by,\ y \mapsto cx + dy$ (affine)
506,508	$x \mapsto ax,\ y \mapsto g(y)$
511	$x \mapsto f(x),\ y \mapsto g(y)$
517–20, 523, 513.1, 513.3, 513.4, 513.6, 514, 525	conformal
524	'peculiar'

Table 1. Transformation groups used in some transformations in Chapter XII, 'On the Theory of Transformations, or the Comparison of Related Forms', of [2]

We draw attention to two key aspects of Thompson's examples: (i) the transformations are as simple as possible to achieve what he considers a good enough match (see Table 1); and (ii) the classes of transformations that he considers all forms groups (or pseudo groups), either finite or infinite dimensional. Mostly, he uses conformal transformations, a constraint he is reluctant to give up[3].

In terms of diffeomorphic image matching, Figs. 506, 508 and 511 of [2] are related to the one-dimensional diffeomorphism group, and hence to the Camassa–Holm family of equations [4]; but the groups usually studied in the literature are the full diffeomorphism group (two functions of two variables) and the volume-preserving group (one function of two variables). Conformal maps are defined by two functions of one variable: a drastically reduced dimensionality.

For applications in image registration we therefore suggest to vary the group from which warps are drawn as well as the metric. If a low-dimensional group gives a close match, then it should be preferred over a similar match from a higher-dimensional group; if necessary, local deformations from the full diffeo-

[3] "It is true that, in a mathematical sense, it is not a perfectly satisfactory or perfectly regular deformation, for the system is no longer isogonal; but[. . .] approaches to an isogonal system under certain conditions of friction or constraint." ([2], p. 1064) "[. . .] it will perhaps be noticed that the correspondence is not always quite accurate in small details. It could easily have been made much more accurate by giving a slightly sinuous curvature to certain of the coordinates. But as they stand, the correspondence indicated is very close, and the simplicity of the figures illustrates all the better the general character of the transformation." (ibid., p. 1074).

morphism group can be added later. In this paper we consider the case of conformal transformations.

Although the composition of two conformal maps $\mathbb{R}^2 \to \mathbb{R}^2$ is conformal, it need not be invertible: we need to restrict the domain. The invertible conformal maps $\mathbb{D} \to \mathbb{D}$ do form a group, the disk-preserving Möbius group, but it is only 3 dimensional. We are therefore led to consider the infinite dimensional configuration space $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ of planar conformal embeddings of the closed unit disk $\mathbb{D}$ into the plane. This is not a group, but it is a pseudo group.

2 Derivation of the Weak Geodesic Equation

In this section we give a derivation of the weak form of the governing equation.

Let g denote the Euclidean metric on $\mathbb{R}^2$, i.e., in Cartesian coordinates we have $\mathsf{g} = \mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y$. Then $(\mathbb{D}, \mathsf{g})$ is a compact Riemannian manifold with boundary. The linear space $C^\infty(\mathbb{D}, \mathbb{R}^2)$ of smooth maps $\mathbb{D} \to \mathbb{R}^2$ is a Fréchet space (see [5, Sect. I.1] for details on the Fréchet topology used). The set $\mathrm{Emb}(\mathbb{D}, \mathbb{R}^2)$ of smooth embeddings of the disk into the plane is an open subset of $C^\infty(\mathbb{D}, \mathbb{R}^2)$ and has the structure of a Fréchet manifold (see [5, Sect. I.4.1]). The subspace $C^\infty_{\mathrm{c}}(\mathbb{D}, \mathbb{R}^2) = \{\phi \in C^\infty(\mathbb{D}, \mathbb{R}^2); \phi^* \mathsf{g} = F\mathsf{g}, F \in \mathcal{F}(\mathbb{D})\}$ of maps preserving the metric up to a scalar function is topologically closed in $C^\infty(\mathbb{D}, \mathbb{R}^2)$. Furthermore, it holds that the set of conformal embeddings $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2) = C^\infty_{\mathrm{c}}(\mathbb{D}, \mathbb{R}^2) \cap \mathrm{Emb}(\mathbb{D}, \mathbb{R}^2)$ is a Fréchet submanifold of $\mathrm{Emb}(\mathbb{D}, \mathbb{R}^2)$.

For standard planar template matching equations, one introduces a positive-definite quadratic Lagrangian function (corresponding to a weak Riemannian structure) on the infinite dimensional manifold $\mathrm{Diff}(\mathbb{D}) = \mathrm{Emb}(\mathbb{D}, \mathbb{D})$. However, as mentioned above, the set $\mathrm{Diff}_{\mathrm{c}}(\mathbb{D})$ of disk preserving conformal maps is small, so we consider instead $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ as configuration manifold. The Riemann mapping theorem asserts that $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ contains a rich set of maps: for any simply connected domain $\mathsf{U} \subset \mathbb{R}^2$ we can find $\varphi \in \mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ such that $\varphi(\mathbb{D}) = \mathsf{U}$, and φ is then unique up to the disk-preserving Möbius transformations. That is, φ must be an element of a unique co-set $[\varphi] \in \mathrm{Con}(\mathbb{D}, \mathbb{R}^2)/\mathrm{Diff}_{\mathrm{c}}(\mathbb{D})$, where $\mathrm{Diff}_{\mathrm{c}}(\mathbb{D})$ acts on $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ by composition from the right. Thus, it holds that $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)/\mathrm{Diff}_{\mathrm{c}}(\mathbb{D})$ is equivalent to the well known *shape space* $\mathrm{Emb}(\mathbb{D}, \mathbb{R}^2)/\mathrm{Diff}(\mathbb{D})$, so we expect that the equation studied in this paper will be relevant not only for conformal image matching, but also for planar shape matching[4]. Also, as developed in [6], a planar shape may, by conformal mappings, be represented by a *fingerprint* in $\mathrm{Diff}(S^1)$, which suggests that $\mathrm{Diff}(S^1)$ may be used for co-sets.

Since $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ is open in the linear space $C^\infty_{\mathrm{c}}(\mathbb{D}, \mathbb{R}^2)$ its tangent bundle is $T\mathrm{Con}(\mathbb{D}, \mathbb{R}^2) = \mathrm{Con}(\mathbb{D}, \mathbb{R}^2) \times C^\infty_{\mathrm{c}}(\mathbb{D}, \mathbb{R}^2)$. The Lagrangian we are interested in is given by

$$L(\varphi, \dot{\varphi}) = \frac{1}{2} \int_{\varphi(\mathbb{D})} \mathsf{g}(\dot{\varphi} \circ \varphi^{-1}, \dot{\varphi} \circ \varphi^{-1}) \, \mathrm{d}A \tag{1}$$

[4] This is only true in $\mathbb{R}^2$: for higher dimensions the set of conformal embeddings of the unit ball $\mathcal{B}$ is very small, and it does not hold that $\mathrm{Con}(\mathcal{B}, \mathbb{R}^n)/\mathrm{Diff}_{\mathrm{c}}(\mathcal{B})$ and $\mathrm{Emb}(\mathcal{B}, \mathbb{R}^n)/\mathrm{Diff}(\mathcal{B})$ are equivalent.

where $\mathrm{d}A = \mathrm{d}x \wedge \mathrm{d}y$ is the standard Euclidean area form on $\mathbb{R}^2$ associated with g.

Note that φ^{-1} is well-defined as a map $\varphi(\mathbb{D}) \to \mathbb{D}$ and that the Lagrangian is well-defined on the full embedding manifold $\mathrm{Emb}(\mathbb{D}, \mathbb{R}^2)$. It is quadratic in $\dot{\varphi}$ and positive-definite, thus determining a (weak) Riemannian metric. When restricted to the submanifold $\mathrm{Diff}(\mathbb{D})$ of $\mathrm{Emb}(\mathbb{D}, \mathbb{R}^2)$, this metric coincides with the metric used in the L^2 template matching equation (TME). Further, $\mathrm{Diff}_\mathrm{c}(\mathbb{D})$ is a right symmetry group: if $\phi \in \mathrm{Diff}_\mathrm{c}(\mathbb{D})$ then $L(\varphi \circ \phi, \dot{\varphi} \circ \phi) = L(\varphi, \dot{\varphi})$. In turn, this implies that we obtain a *reduced variational principle* with respect to the reduced variable $\xi = \dot{\varphi} \circ \varphi^{-1} \in \mathfrak{X}_\mathrm{c}(\varphi(\mathbb{D}))$, where $\mathfrak{X}_\mathrm{c}(\varphi(\mathbb{D}))$ denotes the set of conformal vector fields (whose flow is conformal). However, we still need to keep track of the correct domain $\varphi(\mathbb{D})$, which corresponds to the "shape space" element in $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)/\mathrm{Diff}_\mathrm{c}(\mathbb{D})$, so the complete set of reduced variables consists of a domain $\mathsf{U} \in \mathcal{R}(\mathbb{R}^2)$ (where $\mathcal{R}(\mathbb{R}^2)$ is the Fréchet manifold of simply connected compact planar regions, see [5, Sect. I.4.3]), and a conformal vector field $\xi \in \mathfrak{X}_\mathrm{c}(\mathsf{U})$ defined on this domain. Thus, the reduced phase space is the Fréchet tangent bundle $\mathfrak{X}(\mathcal{R}(\mathbb{R}^2))$ over $\mathcal{R}(\mathbb{R}^2)$, such that the fibre over U is the vector space $\mathfrak{X}_\mathrm{c}(\mathsf{U})$ of conformal vector fields (see [7] for details). For $\xi, \eta \in \mathfrak{X}(\mathsf{U})$ let $\langle \xi, \eta \rangle_\mathsf{U} := \int_\mathsf{U} \mathsf{g}(\xi, \eta) \mathrm{d}A$. The variational principle then becomes

$$0 = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} \int_0^1 L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\, \mathrm{d}t = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} \int_0^1 \frac{1}{2} \langle \xi_\varepsilon, \xi_\varepsilon \rangle_{\mathsf{U}_\varepsilon}\, \mathrm{d}t \tag{2}$$

where φ_ε is a variation of an extremal curve $t \mapsto \varphi(t)$ in $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ and

$$(\mathsf{U}_\varepsilon, \xi_\varepsilon) := (\varphi_\varepsilon(\mathbb{D}), \dot{\varphi}_\varepsilon \circ \varphi_\varepsilon^{-1}).$$

Since a general variation of $t \to \varphi(t)$ is of the form $\exp(\varepsilon\eta) \circ \varphi$ (where $t \mapsto \eta(t)$ is a path in $\mathfrak{X}_\mathrm{c}(\varphi(\mathbb{D}))$ with vanishing endpoints) direct calculation yields

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} \xi_\varepsilon &= \dot{\eta} + \pounds_\eta \xi \\ \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} \langle \xi, \xi \rangle_{\mathsf{U}_\varepsilon} &= \langle \pounds_\eta \xi, \xi \rangle_\mathsf{U} + \int_\mathsf{U} \mathsf{g}(\xi, \xi)\, \mathrm{div}(\eta) \mathrm{vol}. \end{aligned}$$

where $\pounds_\eta$ denotes the Lie derivative along η. Plugging this into (2) yields the weak form of a planar conformal template matching equation (PCTME) to be studied in this paper:

$$\int_0^1 \langle \xi, \dot{\eta} + 2\pounds_\eta \xi + \mathrm{div}(\eta)\xi \rangle_\mathsf{U}\, \mathrm{d}t = 0 \tag{3}$$

for all variations $t \mapsto \eta(t) \in \mathfrak{X}(\mathsf{U}(t))$, where $t \mapsto \mathsf{U}(t)$ fulfills the "compatibility" equation $\dot{\mathsf{U}} = \xi(\partial\mathsf{U})$ and $\mathsf{U}(0) = \mathbb{D}$, i.e., the domain is transported along the flow generated by ξ. In practice, however, we do not solve for the domain variable U, but instead we immediately solve the reconstruction equation $\dot{\varphi} = \xi \circ \varphi$, and use the fact that $\mathsf{U} = \varphi(\mathbb{D})$.

2.1 The Complex Form

By identifying $\mathbb{R}^2$ with $\mathbb{C}$ the space $\mathfrak{X}_c(\mathsf{U})$ is identified with the space of holomorphic functions in the complex domain $\mathsf{U} \subset \mathbb{C}$. Likewise, $C_c^\infty(\mathbb{D}, \mathbb{R}^2)$ is identified with the space of holomorphic function on $\mathbb{D} \subset \mathbb{C}$, and $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$ with the space of injective holomorphic functions on $\mathbb{D}$. We therefore identify $\varphi, \dot{\varphi}$ as holomorphic functions on $\mathbb{D}$, and φ^{-1}, ξ as holomorphic functions on $\varphi(\mathbb{D})$. The complex derivative of a holomorphic function is denoted by a prime, e.g., ξ' for $z \mapsto \xi'(z)$.

When working with complex entries, it is useful to express the weak form (3) in terms of the Hermitian inner product $\langle\!\langle \xi, \eta \rangle\!\rangle_{\mathsf{U}} = \int_{\mathsf{U}} \xi \overline{\eta} \, \mathrm{d}A$ instead of the real-valued inner product.

Direct calculation yields that $\mathsf{g}(\xi, \mathrm{div}(\eta)\xi) + \mathrm{i}\mathsf{g}(\xi, \mathrm{div}(\mathrm{i}\eta)\xi) = 2\xi\overline{\eta'\xi}$. Using this and the fact that $\langle\!\langle \xi, \eta \rangle\!\rangle_{\varphi(\mathbb{D})} = \langle \xi, \eta \rangle_{\varphi(\mathbb{D})} + \mathrm{i}\langle \xi, \mathrm{i}\eta \rangle_{\varphi(\mathbb{D})}$ we obtain the complex weak form of the PCTME (3) as (where we write $\varphi(\mathbb{D})$ instead of U to indicate that we choose to simultaneously solve the reconstruction equation $\dot{\varphi} = \xi \circ \varphi$, instead of only solving for U)

$$\int_0^1 \langle\!\langle \xi, \dot{\eta} + 4\eta'\xi - 2\xi'\eta \rangle\!\rangle_{\varphi(\mathbb{D})} \, \mathrm{d}t = 0. \tag{4}$$

We investigate a sheet ansatz for this equation in Section 4, leading to a numerical discretization, but first we consider a set of special solutions to this equation, by identifying a totally geodesic submanifold of $\mathrm{Con}(\mathbb{D}, \mathbb{R}^2)$.

3 Totally Geodesic Submanifolds

Recall that a submanifold $N \subset M$ of a Riemannian manifold (M, g) is *totally geodesic with respect to* (M, g) if geodesics in N (with respect to g restricted to N) are also geodesics in M. For a thorough treatment of totally geodesic subgroups of $\mathrm{Diff}(M)$ (with respect to various metrics), see [8].

Consider the Fréchet submanifold of linear conformal transformations

$$\mathrm{Lin}(\mathbb{D}, \mathbb{R}^2) = \left\{ \varphi \in \mathrm{Con}(\mathbb{D}, \mathbb{R}^2); \varphi(z) = cz, c \in \mathbb{C} \right\}.$$

Proposition 1. $\mathrm{Lin}(\mathbb{D}, \mathbb{C})$ *is totally geodesic in* $\mathrm{Con}(\mathbb{D}, \mathbb{C})$.

Proof. If $t \mapsto \varphi(t)$ is a path in $\mathrm{Lin}(\mathbb{D}, \mathbb{C})$, i.e., $\varphi(z) = cz$ with $c \in \mathbb{C}$, then $\xi = \dot{\varphi} \circ \varphi^{-1}$ is of the form $\xi(z) = az$ with $a \in \mathbb{C}$. Now, let $t \mapsto (\varphi, \xi)$ fulfill the variational equation (4) for each variation of the form $\eta(z) = bz$ with $b \in \mathbb{C}$. We need to show that $t \mapsto (\varphi, \xi)$ then fulfills the equation for any variation of the form $\eta(z) = ez^k$ (since the monomials span the space of holomorphic functions). Thus,

$$\begin{aligned} \langle\!\langle \xi, \dot{\eta} + 4\eta'\xi - 2\xi'\eta \rangle\!\rangle_{\varphi(\mathbb{D})} &= \langle\!\langle \varphi' \cdot \xi \circ \varphi), \varphi' \cdot (\dot{\eta} + 4\eta'\xi - 2\xi'\eta) \circ \varphi \rangle\!\rangle_{\mathbb{D}} \\ &= |c|^4 \langle\!\langle az, \dot{b}z^k + 4kbaz^k - 2abz^k \rangle\!\rangle_{\mathbb{D}}. \end{aligned} \tag{5}$$

where in the first line we use the conformal change of variables formula for integrals. Now, since the monomials are orthogonal with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{D}}$ the expression vanish whenever $k \neq 1$, which concludes the proof. □

From equation (5) we may derive a differential equation for the totally geodesic solutions in $\mathrm{Lin}(\mathbb{D},\mathbb{C})$ in term of the variables (c,a) corresponding to $\varphi(z)=cz$ and $\xi(z)=az$. Indeed, choosing $k=1$ and plugging equation (5) into the variational equation (4) we obtain

$$\begin{aligned}0&=\int_0^1 |c|^4(a\dot{\bar{b}}+2|a|^2\bar{b})\,\mathrm{d}t=\int_0^1\Big(\frac{\mathrm{d}}{\mathrm{d}t}a\bar{b}|c|^4-\bar{b}\big(\frac{\mathrm{d}}{\mathrm{d}t}a|c|^4-2|a|^2|c|^4\big)\Big)\,\mathrm{d}t\\&=-\int_0^1\bar{b}\big(\dot{a}|c|^4+2a\dot{c}\bar{c}|c|^2+2a\dot{\bar{c}}c|c|^2-2|a|^2|c|^4\big)\\&=-\int_0^1\bar{b}|c|^4\big(\dot{a}+2a\big)\end{aligned}$$

where we have used that b vanish at the endpoints, and in the last line we use that $\dot{c}=ac$, which follows since $\dot{\varphi}=\xi\circ\varphi$. Hence, the governing equations for the totally geodesic solutions are

$$\begin{aligned}\dot{c}&=ac\\\dot{a}&=-2a^2\end{aligned}\tag{6}$$

where $c(0)=1$ (since $\varphi(z)=z$ initially). We can see that if $a(0)$ is real, then both a and c stay real, so that the smaller submanifold of pure scalings is also totally geodesic. However, pure rotations are *not* totally geodesic.

By using again the weak form (4) of the governing equation one can further show that the submanifold of translations is *not* totally geodesic in $\mathrm{Con}(\mathbb{D},\mathbb{C})$. Nor is the submanifold of affine conformal transformations. The result is somewhat surprising, given that these types of basic transformations are totally geodesic for the L^2–TME and H^1_α–TME[5], if boundary conditions are set up to allow them [8]. In fact, from a template matching point of view, where basic transformations should preferably be totally geodesic, our result indicates that the metric we have picked is not ideal.

4 Sheet Ansatz

The L^2–TME and H^1_α–TME are known to admit solutions corresponding to totally geodesic submanifolds described by momentum particles or sheets [9]. These special solutions can be used for structure preserving numerical discretization [10,11]. Since the ansatz is non-smooth (the momentum is a traveling peakon), it cannot describe solutions to the conformal equation studied here. However, in this section we show that it is possible to express solutions to the PCTME (4) with a different sheet ansatz, using a reproducing kernel.

Let $\mathcal{O}\subset\mathbb{C}$ be an open bounded complex region. Then $A^2(\mathcal{O})=\{f\in L^2(\mathcal{O});\, f \text{ is holomorphic}\}$ is a Hilbert space with the inner product $\langle\!\langle\cdot,\cdot\rangle\!\rangle_{\mathcal{O}}$ (called

[5] Also called the *EPDiff equation* or the *averaged template matching equation*.

a *Bergman space*, see [12, Ch. 1]). For any $z \in \mathcal{O}$ it holds that point-wise evaluation $E_z : \xi \mapsto \xi(z)$ is continuous with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathcal{O}}$. Thus, by the Riesz representation theorem there exists a unique $k_z \in A^2(\mathcal{O})$ such that $f(z) = \langle\!\langle f, k_z \rangle\!\rangle_{\mathcal{O}}$ for every $f \in A^2(\mathcal{O})$. Let $\mathsf{U} = \overline{\mathcal{O}}$ be the closure of $\mathcal{O}$. Since $\mathfrak{X}_{\mathrm{c}}(\mathsf{U}) \subset A^2(\mathcal{O})$ it clearly holds that $\xi(z) = \langle\!\langle \xi, k_z \rangle\!\rangle_{\mathsf{U}}$ for all $\xi \in \mathfrak{X}_{\mathrm{c}}(\mathsf{U})$ (however, notice that it does *not* hold that $k_z \in \mathfrak{X}_{\mathrm{c}}(\mathsf{U})$). The function defined by $K_{\mathsf{U}}(z, w) = \overline{k_z(w)}$ is called the *reproducing kernel* (or sometimes the *Bergman kernel function*) for the domain U. On the interior of U this function is analytic in z and anti-analytic in w, with $K(z, w) = \overline{K(w, z)}$. Furthermore, if $\varphi \in \mathrm{Con}(\mathbb{D}, \mathbb{C})$, then

$$K_{\varphi(\mathbb{D})}(z, w) = \frac{K_{\mathbb{D}}(\varphi^{-1}(z), \varphi^{-1}(w))}{\varphi'(\varphi^{-1}(z))\overline{\varphi'(\varphi^{-1}(w))}} \tag{7}$$

which follows by a change of variables in the integral over $\varphi(\mathbb{D})$. When working with kernels in practice, e.g., for numerical purposes, this formula is very useful since the kernel function for the disk is known to be

$$K_{\mathbb{D}}(z, w) = \frac{1}{\pi} \frac{1}{(1 - z\overline{w})^2} \quad .$$

Now, consider the following ansatz:

$$\xi(z) = \int_0^{2\pi} p_\varphi(\rho \mathrm{e}^{\mathrm{i}s}) K_{\varphi(\mathbb{D})}\big(z, \varphi(\rho \mathrm{e}^{\mathrm{i}s})\big) \mathrm{d}s \tag{8}$$

where $p_\varphi \in C_{\mathrm{c}}^\infty(\mathbb{D}, \mathbb{C})$ and $0 < \rho < 1$ is a fixed constant. We introduce the periodic functions $\gamma, p_\gamma : S^1 \to \mathbb{C}$ defined by $\gamma(s) = \varphi(\rho \mathrm{e}^{\mathrm{i}s})$ and $p_\gamma(s) = p_\varphi(\rho \mathrm{e}^{\mathrm{i}s})$, so that $\xi(z) = \int_0^{2\pi} p_\gamma K_{\varphi(\mathbb{D})}(z, \gamma) \mathrm{d}s$. One may think of $\gamma(s)$ as just a notational shortcut, but it has other significance: by expanding φ as a Taylor series with coefficients $(c_k)_{k=0}^\infty$, we see that the periodic curve $\gamma(s)$ only has positive Fourier coefficients, which are given by $(\rho^k c_k)_{k=0}^\infty$. An equivalent statement is to say that the Fourier coefficients of $\gamma(s)$ are the Taylor coefficients of $z \mapsto \varphi(\rho z)$. Exactly the same relation holds between p_γ and p_φ. This observation can be used in the numerical discretization to accelerate the evaluation of the right-hand side using fast Fourier transformations.

Plugging the ansatz (8) into the weak equation (4) yields

$$\begin{aligned}
0 &= \int_0^1 \int_0^{2\pi} \overline{p_\gamma} \langle\!\langle \dot{\eta} + 4\eta'\xi - 2\xi'\eta, K_{\varphi(\mathbb{D})}(\cdot, \gamma) \rangle\!\rangle_{\varphi(\mathbb{D})} \, \mathrm{d}s \, \mathrm{d}t \\
&= \int_0^1 \int_0^{2\pi} \overline{p_\gamma} \Big(\dot{\eta}(\gamma) + 4\eta'(\gamma)\xi(\gamma) - 2\xi'(\gamma)\eta(\gamma) \Big) \, \mathrm{d}s \, \mathrm{d}t \\
&= \int_0^1 \int_0^{2\pi} \Big(\eta(\gamma) \Big(-\overline{\dot{p}_\gamma} - 2\overline{p_\gamma}\xi'(\gamma) \Big) + \overline{p_\gamma}\eta'(\gamma) \Big(4\xi(\gamma) - \dot{\gamma} \Big) \Big) \, \mathrm{d}s \, \mathrm{d}t
\end{aligned}$$

where we use the evaluation property of the kernel in the second line and the fact that $t \mapsto \eta(t)$ vanish at the endpoints in the third. Next, using that $\dot{\gamma}(s) =$

$\dot{\varphi}(\rho e^{is}) = \xi(\varphi(\rho e^{is})) = \xi(\gamma(s))$, we get

$$\begin{aligned} 0 &= \int_0^1 \int_0^{2\pi} \left(\eta(\gamma) \Big(-\overline{\dot{p}_\gamma} - 2\overline{p_\gamma}\xi'(\gamma) \Big) + \overline{p_\gamma}\eta'(\gamma)\, 3\xi(\gamma) \right) \mathrm{d}s\, \mathrm{d}t \\ &= \int_0^1 \int_0^{2\pi} \left(\eta(\gamma) \Big(-\overline{\dot{p}_\gamma} - 2\overline{p_\gamma}\xi'(\gamma) \Big) + \eta'(\gamma)\gamma' \left(\frac{3\overline{p_\gamma}\xi(\gamma)}{\gamma'} \right) \right) \mathrm{d}s\, \mathrm{d}t \\ &= \int_0^1 \int_0^{2\pi} \left(\eta(\gamma) \left(-\overline{\dot{p}_\gamma} - 2\overline{p_\gamma}\xi'(\gamma) \right) - \eta(\gamma) \frac{\partial}{\partial s} \left(\frac{3\overline{p_\gamma}\xi(\gamma)}{\gamma'} \right) \right) \mathrm{d}s\, \mathrm{d}t \\ &= -\int_0^1 \int_0^{2\pi} \eta(\gamma) \left(\overline{\dot{p}_\gamma} - 2\xi(\gamma) \frac{\partial}{\partial s} \left(\frac{\overline{p_\gamma}}{\gamma'} \right) + 5 \frac{\partial}{\partial s} \left(\frac{\overline{p_\gamma}\xi(\gamma)}{\gamma'} \right) \right) \mathrm{d}s\, \mathrm{d}t \end{aligned} \tag{9}$$

where γ' denotes the derivative of γ with respect to s. Thus, by inserting the ansatz (8) we obtain a weak formulation of the PCTME, equivalent to (4), in terms of the variables γ and p_γ. This form involves an integral operator, since ξ is related to γ, p_γ by the ansatz (8).

In contrast to the L^2–TME and H^1_α–TME, *any* solution to the PCTME can be represented by the sheet ansatz, since any φ can be reconstructed from γ by the Fourier transformation.

A nice feature of the weak formulation (9) is that the test function η is isolated, i.e., without derivatives. This allows us to use the fundamental lemma of calculus of variations to get rid of the outermost time integral (so we get a condition at each fixed point in time). However, we do not immediately obtain a strong formulation, because the test function η must be holomorphic, so we cannot use the fundamental lemma of calculus of variations for the inner integral without introducing a projection operator. Since the trigonometric monomials e^{iks} are orthogonal with respect to the s–integral, the projection is given by neglecting negative Fourier coefficients. Hence, let $\chi_+ : \mathbb{Z} \to \mathbb{R}$ be the sequence

$$\chi_+(k) = \begin{cases} 1 \text{ if } k \geq 0 \\ 0 \text{ if } k < 0 \end{cases}$$

and let $\mathcal{F}$ denote the Fourier operator taking a 2π–periodic function to its Fourier series. Then the projection operator is given by $\mathcal{P} = \mathcal{F}^{-1}\chi_+\mathcal{F}$ (where χ_+ is acing by element-wise multiplication). We now obtain a strong integral formulation of the PCTME in terms of γ and p_γ as

$$\begin{aligned} \dot{\gamma} &= \xi(\gamma) \\ \dot{p}_\gamma &= \mathcal{P}\left(2\overline{\xi(\gamma)} \frac{\partial}{\partial s} \frac{p_\gamma}{\overline{\gamma'}} - 5 \frac{\partial}{\partial s} \frac{p_\gamma \overline{\xi(\gamma)}}{\overline{\gamma'}} \right) . \end{aligned} \tag{10}$$

5 Numerical Discretization

In this section we describe a method for numerical discretization of equations (10). The idea is to represent the dynamic variables $\gamma(s)$ and $p_\gamma(s)$ by truncated positive Fourier series, or, equivalently, truncated Taylor series of $z \mapsto \varphi(\rho z)$ and

$z \mapsto p_\varphi(\rho z)$. Thus, our finite set of dynamic variables are $\boldsymbol{c} = (c_k)_{k=0}^{n-1} \in \mathbb{C}^n$ and $\boldsymbol{p_c} = (p_k)_{k=0}^{n-1} \in \mathbb{C}^n$ such that

$$\gamma(s) = \varphi(\rho \mathrm{e}^{\mathrm{i}s}) = \sum_{k=0}^{n-1} c_k \mathrm{e}^{\mathrm{i}s}, \quad p_\gamma(s) = p_\varphi(\rho \mathrm{e}^{\mathrm{i}s}) = \sum_{k=0}^{n-1} p_k \mathrm{e}^{\mathrm{i}s}. \tag{11}$$

For the discretization we insert these into the right-hand side of equation (10) and compute the first n positive Fourier coefficients, which then gives an ordinary differential equation for $(\boldsymbol{c}, \boldsymbol{p_c}) \in \mathbb{C}^{2n}$. We now describe how this discretized right-hand side is computed efficiently.

1. Compute $\boldsymbol{\gamma} = (\gamma(0), \gamma(\pi/n), \dots, \gamma(\pi(2n-1)/n))$ from $\boldsymbol{c}$. (Notice that $\boldsymbol{\gamma}$ has length $2n$.) Since $\gamma(s) = \varphi(\rho \mathrm{e}^{\mathrm{i}s})$ and $\boldsymbol{c}$ are the Taylor coefficients of $z \mapsto \varphi(\rho z)$, we obtain $\boldsymbol{\gamma}$ from the inverse FFT as $\boldsymbol{\gamma} = 2n \operatorname{IFFT}(\boldsymbol{c}^+)$, where $\boldsymbol{c}^+ = (\boldsymbol{c}, 0, \dots, 0)$ is zero-padded to get the same length as $\boldsymbol{\gamma}$ (since in the FFT the last half of the vector corresponds to negative frequencies).
2. Compute $\boldsymbol{p_\gamma} = (p_\gamma(0), \dots, p_\gamma(\pi(2n-1)/n))$ from $\boldsymbol{p_c}$. By the same argument we have $\boldsymbol{p_\gamma} = 2n \operatorname{IFFT}(\boldsymbol{p_c})$.
3. Compute $\boldsymbol{\gamma}' = (\gamma'(0), \gamma'(2\pi/n), \dots, \gamma'(2\pi(n-1)/n))$. Since the Fourier coefficients of $\gamma(s)$ are given by $\boldsymbol{c}$, we have $\boldsymbol{\gamma}' = 2n \operatorname{IFFT}(\mathrm{i}\boldsymbol{k}^+ \cdot \boldsymbol{c}^+)$ where $\boldsymbol{k} = (0, 1, \dots, n-1)$ and $\boldsymbol{k}^+ \cdot \boldsymbol{c}^+$ denotes element-wise multiplication.
4. Compute $\boldsymbol{\xi} = (\xi(\boldsymbol{\gamma}_0), \dots, \xi(\boldsymbol{\gamma}_{2n-1}))$. From the ansatz (8) and the equation for the Bergman kernel (7), we get

$$\xi(\gamma(s)) = \int_0^{2\pi} \frac{p_\gamma(\sigma) K_{\mathbb{D}}(\rho \mathrm{e}^{\mathrm{i}s}, \rho \mathrm{e}^{\mathrm{i}\sigma})}{\varphi'(\rho \mathrm{e}^{\mathrm{i}s}) \overline{\varphi'(\rho \mathrm{e}^{\mathrm{i}\sigma})}} \mathrm{d}\sigma = \frac{1}{\pi \gamma'(s)} \int_0^{2\pi} \frac{p_\gamma(\sigma) G(\rho^2 \mathrm{e}^{\mathrm{i}(s-\sigma)})}{\overline{\gamma'(\sigma)}}$$

   where $G(z) = z/(1-z)^2$. Now, set $f(s) = p_\gamma(s)/\overline{\gamma'(s)}$ and $g(s) = G(\rho^2 \mathrm{e}^{\mathrm{i}s})$. Then

$$\xi(\gamma(s)) = \frac{1}{\pi \gamma'(s)} \int_0^{2\pi} f(\sigma) g(s-\sigma) \mathrm{d}\sigma = \frac{(f * g)(s)}{\pi \gamma'(s)}.$$

   Thus, using that a convolution becomes element-wise multiplication in the Fourier domain, we get $\boldsymbol{\xi} = \frac{2n}{\pi} \operatorname{IFFT}(\hat{\boldsymbol{f}} \cdot \hat{\boldsymbol{g}})/\boldsymbol{\gamma}'$ where $\hat{\boldsymbol{g}} = (0, \rho^2, \dots, \rho^{2(n-1)}(n-1), 0, \dots, 0)$, which follows by computing the Taylor coefficients of $G(z)$, and $\hat{\boldsymbol{f}} = \operatorname{FFT}(\boldsymbol{p_\gamma}/\overline{\boldsymbol{\gamma}'})/2n$ (divisions are carried out element-wise).
5. Compute $\dot{\boldsymbol{c}}$. Using again the correspondence between Taylor and Fourier coefficients, we get $\dot{\boldsymbol{c}} = \operatorname{FFT}(\boldsymbol{\xi})^-/2n$, where the $\operatorname{FFT}(\boldsymbol{\xi})^-$ means that we only keep the first n elements. Thus, we have now computed the first half of the right-hand side.
6. Compute $\dot{\boldsymbol{p}}_{\boldsymbol{c}}$. From the second equation in (10) we get

$$\dot{\boldsymbol{p}}_{\boldsymbol{c}} = \operatorname{FFT}(2\overline{\boldsymbol{\xi}} \cdot \operatorname{IFFT}(\mathrm{i}\boldsymbol{k}^+ \cdot \hat{\boldsymbol{f}}))^- - 5\mathrm{i}\boldsymbol{k} \cdot \operatorname{FFT}(\overline{\boldsymbol{\xi}} \cdot \boldsymbol{p_\gamma}/\overline{\boldsymbol{\gamma}'})^-/2n.$$

   Note that the projection (corresponding to the operator $\mathcal{P}$ in equation (10)) occurs in the above computation since we only keep elements corresponding to positive frequencies.

We have now computed the full right-hand side $(\dot{\boldsymbol{c}}, \dot{\boldsymbol{p}}_{\boldsymbol{c}})$. By using a time-stepping method for ODEs (e.g., a Runge-Kutta method), a numerical method for equation (10) is obtained. Each evaluation of the right-hand side requires element-wise operations and 9 FFTs on vectors of length $2n$, which leads to complexity $\mathcal{O}(n \log(2n))$.

6 Experimental Results

In this section we study the dynamical behaviour of some solutions to the PCTME. First, we look at the solutions corresponding to the totally geodesic submanifold derived in Section 3. Thereafter, we use the numerical method derived in Section 5 to study other solutions. In particular, we study the spectrum of small perturbations of a totally geodesic solution.

6.1 Totally Geodesic Solutions

Let $\varphi_0(z) = z$ be the identity maps, and let $\varphi_1(z) = c_1 z$ where $c_1 \in \mathbb{C}\backslash\{0\}$. Since both φ_0 and φ_1 belong to the totally geodesic submanifold $\mathrm{Lin}(\mathbb{D}, \mathbb{C})$, the geodesic from φ_0 to φ_1 stays in $\mathrm{Lin}(\mathbb{D}, \mathbb{C})$. In Fig. 1 the geodesics for $c_1 = 0.2$ (scaling of the disk) and $c_1 = \mathrm{e}^{1.6\pi\mathrm{i}}$ (rotation of the disk) are shown. The solutions confirm what we earlier noticed, that pure scalings are totally geodesic (the curve $\varphi(t)$ remains a scaling for each t) whereas the geodesic corresponding to the rotated disk does not stay a pure rotation, but contains scaling along its path.

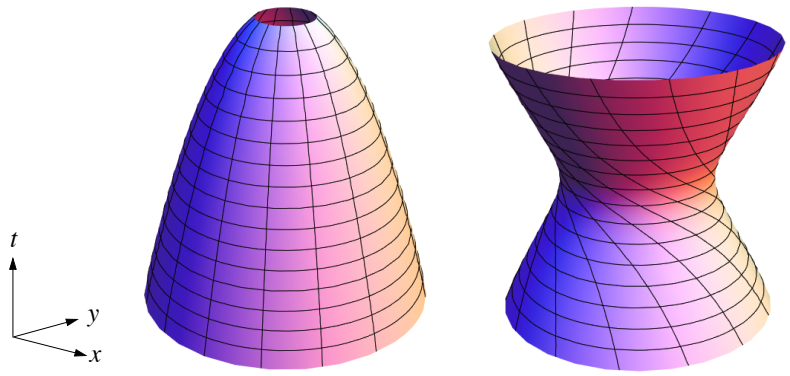


Fig. 1. Geodesic curve from $\varphi_0(z) = z$ to $\varphi_1(z) = c_1 z$ for different values of c_0. The mesh lines show how the unit circle evolves. Notice that the scaling geodesic stays a scaling (left figure), whereas the rotation geodesic picks up some scaling during its time evolution (right figure).

6.2 Development of Cusps in Finite Time

The L^2–TME is known to develop singularities in finite time. Originally, we expected solutions to the PCTME be more "well behaved" due to the holomorphic constraint. However, experiments with the numerical method derived in Section 5 indicates that the same phenomenon still occurs despite of the constraint. For example, Fig. 2 shows an experiment with initial data of the form $\dot{\varphi}_0(z) = a_2 z^2$. A cusp develops in the solution, which leads to a breakdown of the dynamics.

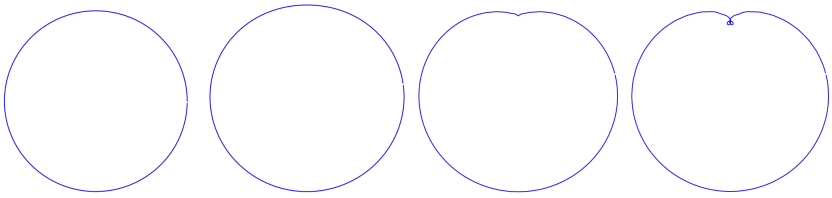


Fig. 2. Time evolution of the PCTME with initial conditions of the form $\varphi_0(z) = z$ and $\dot{\varphi}_0(z) = a_2 z^2$. From left to right the plots show the development of a cusp. Thus, the geodesic curve ceases to exist in finite time.

6.3 Spectral Behaviour

In Fig. 3 the time evolution of the absolute values of $\boldsymbol{p_c}$, computed with the numerical method in Section 5, are shown in two plots. Dark colours corresponds to low order coefficients, and light colours to high order coefficients. In the top plot, initial data corresponding to a totally geodesic solution in $\mathrm{Lin}(\mathbb{D}, \mathbb{C})$ are used. In the bottom plot, this initial data is perturbed $\dot{\varphi}_0(z) = a_1 z + \delta(z)$, where $\delta(z)$ is a small perturbation of the first five Taylor coefficients.

Notice in both plots that the initially suppressed coefficients grow exponentially fast, which indicates that the totally geodesic solutions are not stable. Also, notice that the growth rate increases with the order, which indicates that the PCTME is ill-conditioned, as is also the case for the L^2–TME. Again, our investigation indicates that the dynamics of the PCTME is similar to that of the L^2–TME. We plan to carry out a more thorough investigation of the spectral behaviour, by using both analytical and numerical techniques.

7 Conclusions

In this paper we have studied a geodesic equation on the manifold of planar conformal embeddings. We showed that that the equations have a totally geodesic submanifold corresponding to linear conformal maps. We also showed that the submanifold of affine conformal maps is not totally geodesic. Numerical experiments indicates that the dynamic behaviour of the equation is similar to that of the L^2–TME.

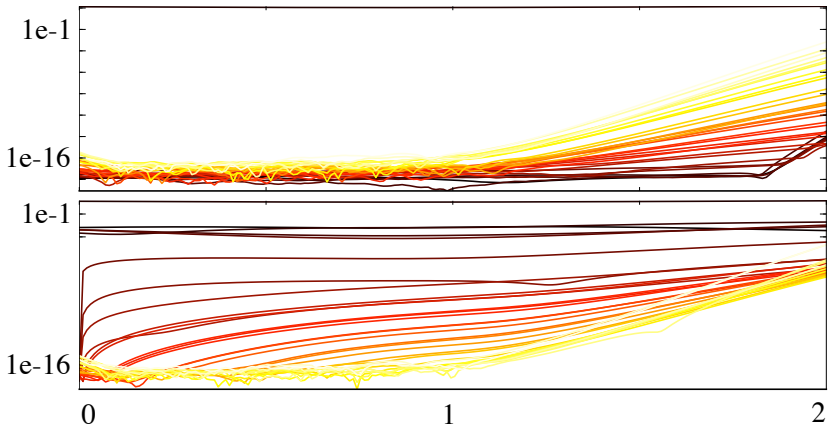


Fig. 3. Time evolution of the absolute values of $\boldsymbol{p_c}$ for a totally geodesic solution computed with the numerical method. *(Top):* without perturbation, *(bottom):* a random perturbation of size $\sim 10^{-4}$ of initial data for the first five coefficients.

In future work we will study the spectral behaviour more thoroughly. Also, as an approach for obtaining more well behaved dynamics, we will consider other metrics, in particular H^1_α. We will also look into more advanced numerical techniques for solving the equations.

References

1. Younes, L.: Shapes and Diffeomorphisms. Applied Mathematical Sciences. Springer-Verlag (2010)
2. Thompson, D.: On Growth and Form. Cambridge University Press (1942)
3. Wallace, A.: D'Arcy Thompson and the theory of transformations. Nature Reviews Genetics **7** (2006) 401–406
4. Camassa, R., Holm, D.D.: An integrable shallow water equation with peaked solitons. Phys. Rev. Lett. **71** (1993) 1661–1664
5. Hamilton, R.S.: The inverse function theorem of Nash and Moser. Bull. Amer. Math. Soc. (N.S.) **7** (1982) 65–222
6. Sharon, E., Mumford, D.: 2d-shape analysis using conformal mapping. Int. J. Comput. Vis. **70** (2006) 55–75 10.1007/s11263-006-6121-z.
7. Marsland, S., McLachlan, R.I., Modin, K., Perlmutter, M.: Reduced geodesic equation for planar conformal embeddings. In preparation (2011)
8. Modin, K., Perlmutter, M., Marsland, S., McLachlan, R.I.: On Euler-Arnold equations and totally geodesic subgroups. J. Geom. Phys. **61** (2011) 1446–1461
9. Holm, D.D., Marsden, J.E.: Momentum maps and measure-valued solutions (peakons, filaments, and sheets) for the EPDiff equation. In: The breadth of symplectic and Poisson geometry. Volume 232 of Progr. Math. Birkhäuser Boston, Boston, MA (2005) 203–235
10. McLachlan, R.I., Marsland, S.: N-particle dynamics of the Euler equations for planar diffeomorphisms. Dyn. Syst. **22** (2007) 269–290
11. Chertock, A., Toit, P., Marsden, J.E.: Integration of the EPDiff equation by particle methods. (2009)
12. Duren, P., Schuster, A.: Bergman Spaces. Volume 100 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI (2004)

Geodesics, Parallel Transport & One-parameter Subgroups for Diffeomorphic Image Registration

Marco Lorenzi[1,2] and Xavier Pennec[1]

[1] Project Team Asclepios, INRIA Sophia Antipolis, France
[2] LENITEM, IRCCS San Giovanni di Dio, Fatebenefratelli, Brescia, Italy

Abstract. The aim of computational anatomy is to develop models for understanding the physiology of organs and tissues. The diffeomorphic non-rigid registration is a validated instrument for the detection of anatomical changes on medical images and is based on a rich mathematical background. For instance, the "large deformation diffeomorphic metric mapping" framework defines a Riemannian setting by providing an opportune right invariant metric on the tangent space, and solves the registration problem by computing geodesics parametrized by time-varying velocity fields. In alternative, stationary velocity fields have been proposed for the diffeomorphic registration based on the one-parameter subgroups from Lie groups theory. In spite of the higher computational efficiency, the geometric setting of the latter method is more vague, especially regarding the relationship between one-parameter subgroups and geodesics. In this study, we present the relevant properties of the Lie groups for the definition of geometrical properties within the one-parameter subgroups parametrization, and we define the geometric structure for computing geodesics and for parallel transporting. The theoretical results are applied to the image registration context, and discussed in light of the practical computational problems.

1 Introduction

Main objective of the computational anatomy is to develop suitable statistical models on several subjects for understanding the physiology of organs and tissues. In particular, the longitudinal observations from time series of images are an important source of information for understanding the developmental processes and the dynamics of pathologies. Thus, a reliable method for comparing different longitudinal trajectories is required, in order to develop population-based longitudinal models.

Non-rigid registration is a validated instrument for the detection of anatomical changes on medical images, and it has been widely applied on different clinical contexts for the definition of population-based anatomical atlases ([14],[9],[3]). However, in case of longitudinal data, the optimal method for comparing deformation trajectories across different subjects is still under discussion. In fact, the methods for integrating the subtle inter-subject changes into the group-wise analysis have an important impact on the accuracy and reliability of the subsequent

results. The aim is to preserve as much as possible the biological informations carried on by the different subjects, while allowing a precise comparison in a common geometric space.

Among the different techniques proposed for the comparison of longitudinal trajectories ([12],[2],[4]), *parallel transport* represents a promising method which relies on a solid mathematical background. Basically, it consists in transporting the infinitesimal deformation vector across different points by preserving its properties with respect to the space geometry, such as the parallelism.

The parallel transport has been introduced for the first time in medical imaging with the LDDMM setting [16]. LDDMM solves the image registration problem by using a Riemannian framework in which the deformations are parametrized as *diffeomorphisms* living in a suitable space, once provided an opportune right-invariant metric [10]. The registration problem is solved by computing the diffeomorphisms lying on the *geodesics* of the space parametrized by *time-varying* velocity fields under the Riemannian exponential. The setting allows the computation of the parallel transport along geodesics at the cost of a computationally intensive scheme, and this limitation prevents the application on high resolution images or large datasets.

A more efficient solution to the image diffeomorphic registration problem was introduced by the stationary velocity field (SVF) setting [1]. In this case, the diffeomorphisms are parametrized as one-parameter subgroups by *stationary* velocity fields through the Lie group exponential. This restriction allows an efficient numerical scheme for the computation but it does not directly rely on any geometric assumption on the underlying space. This implies that some important mathematical properties are not guaranteed, for instance whether the one-parameter subgroups are still geodesics or if the space is metrically complete. In spite of this lack of knowledge, the framework was found very efficient and reliable in many applications in different contexts ([8],[7],[13]) and, in [6], a framework based on the Schild's Ladder has been proposed for the evaluation of the parallel transport with the SVF.

In this paper, we investigate the relationship between Lie groups and Riemannian geometry and we highlight many interesting properties that might provide the SVF setting with part of the geometrical solidity required. In Section 2 we present the relevant properties of the Lie groups and the relationship with the Riemannian setting for the definition of the geodesics and the parallel transport. In Section 3, the results are introduced and discussed for the image registration context, while in Section 4 we show how these theoretical insights provide a clean, precise, and numerically efficient solution for the parallel transport of deformation trajectories on time series of images.

2 Lie Group and covariant differentiation

This section will recall the conceptual basis for the definition of the parallel transport along the one-parameter subgroups.

Let $\mathbb{G}$ an arbitrary finite dimensional Lie group and let $\mathfrak{g}$ the associated Lie algebra defined here with the tangent space at the identity $T_{id}\mathbb{G}$. We define the left translation L_a as the mapping $La : g \mapsto ag$, and we say that a vector field $\mathbf{X} \in T(\mathbb{G})$ is left invariant if $DL_a(\mathbf{X})_b = (\mathbf{X})_{ab}$.
There is a one-to-one correspondence between left-invariant vector fields and elements of the Lie algebra $\mathfrak{g}$, which associates to each $\mathbf{X} \in \mathfrak{g}$ the vector field defined as $\tilde{\mathbf{X}}(g) = DL_g\mathbf{X}$. The left-invariant vector fields are *complete* and their associated flow φ_t is such that $\varphi_t(g) = g\varphi_t(id)$. The association $\mathbf{X} \mapsto \varphi_1(id)$ of $\mathfrak{g}$ into $\mathbb{G}$ is called *Lie group exponential* and denoted by exp. In particular, the map exp defines the *one-parameter subgroup* associated to the vector $\mathbf{X}$ and has the following properties:

- $\varphi_t(id) = \exp(t\mathbf{X})$, for each $t \in \mathbb{R}$
- $\exp((t+s)\mathbf{X}) = \exp(t\mathbf{X})\exp(s\mathbf{X})$, for each $t, s \in \mathbb{R}$

It can be shown that the Lie group exponential is a diffeomorphism from a neighborhood of 0 in $\mathfrak{g}$ to a neighborhood of id in $\mathbf{G}$.

We are going to illustrate the transport of vectors along the exponential path $\exp(t\mathbf{X})$, and in particular the analogies with the classical Riemannian parallel transport defined for *geodesics*.

An *affine connection* on $\mathbb{G}$ is an operator which assigns to each $X \in T(\mathbb{G})$ a linear mapping $\nabla_{\mathbf{X}} : T(\mathbb{G}) \to T(\mathbb{G})$ such that

$$\nabla_{f\mathbf{X}+g\mathbf{Y}} = f\nabla_{\mathbf{X}} + g\nabla_{\mathbf{Y}} \tag{1}$$

$$\nabla_{\mathbf{X}}(f\mathbf{Y}) = f\nabla_{\mathbf{X}}(\mathbf{Y}) + (\mathbf{X}f)\mathbf{Y} \tag{2}$$

A vector field $\mathbf{X}$ is *parallel transported* along a curve $\gamma(t)$ if $\nabla_{\dot{\gamma}(t)}\mathbf{X} = 0$ for each t. In particular, a path $\gamma(t)$ on $\mathbf{G}$ is then said *geodesic* if $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$. The definition generalizes the concept of "straight lines", by requiring to the tangent vector of the path to be covariantly constant.
Given a point $p \in \mathbb{G}$ and a vector $\mathbf{X} \in T_p\mathbb{G}$, there exist a unique geodesic $\gamma(t, p, \mathbf{X})$ such that at the instant $t = 0$ passes through p with velocity $\mathbf{X}$. We define therefore the *Riemannian exponential* as the application $exp : \mathbb{G}\times T(\mathbb{G}) \to \mathbb{G}$ given by $exp_p(\mathbf{X}) = \gamma(1, p, \mathbf{X})$.

If, as in the euclidean case, we want to associate to the straight lines the property of minimizing the distance between points, we need to provide the group $\mathbb{G}$ with a Riemannian manifold structure, i.e. with a metric operator g on the tangent space. In this case there is a unique symmetric connection compatible with

the metric in the sense that, for each $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in T(\mathbb{G})$ the following conditions hold:

$$\begin{aligned} \mathbf{X}\, g(\mathbf{Y}, \mathbf{Z}) &= g(\nabla_{\mathbf{X}}\mathbf{Y}, \mathbf{Z}) + g(\mathbf{X}, \nabla_{\mathbf{X}}\mathbf{Z}) && \text{(Compatibility wrt the metric)} \\ \nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} &= [\mathbf{X}, \mathbf{Y}] && \text{(Torsion free)} \end{aligned}$$

With the choice of this special connection, called *Levi-Civita connection*, the corresponding geodesics (Riemannian geodesics) $\gamma(t)$ are the length minimizing path.

2.1 Relationship between Riemannian geodesic and one-parameter subgroups

Given a vector $\mathbf{X}$ on $T_{id}\mathbb{G}$, we can therefore define two curves on $\mathbb{G}$ passing through id and having $\mathbf{X}$ as tangent vector, one given by the Lie group exponential exp and the other given by the Riemannian exponential exp_{Id}. When do they coincide?

The connection ∇ on $\mathbb{G}$ is *left-invariant* if, for each left translation L_a ($a \in \mathbb{G}$), we have $\nabla_{DL_a\mathbf{X}}(DL_a\mathbf{Y}) = DL_a\nabla_{\mathbf{X}}(\mathbf{Y})$.

A left-invariant connection ∇ on a space $\mathbb{G}$ is a *Cartan connection* if, for any element of the Lie algebra $\mathbf{X} \in \mathfrak{g}$, the one-parameter subgroups and the Riemannian geodesics coincide, i.e. $\exp(t\mathbf{X}) = exp(t, id, \mathbf{X})$ [11].
For each left-invariant connection ∇ we can univoquely associate a product α (symmetric bilinear operator) on $T_{id}\mathbb{G}$ given by

$$\alpha(\mathbf{X}, \mathbf{Y}) = \left(\nabla_{\tilde{\mathbf{X}}}\tilde{\mathbf{Y}}\right)_{id}$$

where $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$ are the unique left-invariant vector fields induced by the tangent vectors $\mathbf{X}, \mathbf{Y}$. We note here that a bilinear for can be uniquely decomposed as $\alpha = \alpha' + \alpha''$, where $\alpha' = \frac{1}{2}(\alpha(X,Y) + \alpha(Y,X))$ is commutative, while $\alpha'' = \frac{1}{2}(\alpha(X,Y) - \alpha(Y,X))$ is skew-symmetric.

We deduce that the condition for ∇ to be a Cartan connection is to satisfy $\alpha(\mathbf{X}, \mathbf{X}) = 0$ or, equivalently, to be skew-symmetric, for instance by assigning

$$\alpha(\mathbf{X}, \mathbf{Y}) = \lambda[\mathbf{X}, \mathbf{Y}] \tag{3}$$

In this case, the zero curvature connections are given by $\lambda = 0, 1$ (with torsion $T = -[\mathbf{X}, \mathbf{Y}]$ and $T = [\mathbf{X}, \mathbf{Y}]$ respectively) and are called *left* and *right* Cartan connections.
The choice of $\lambda = \frac{1}{2}$ lead to the *symmetric* (or *mean*) Cartan connection $\nabla_{\mathbf{X}}\mathbf{Y} = \frac{1}{2}[\mathbf{X}, \mathbf{Y}]$, with curvature $C = -\frac{1}{4}[[\mathbf{X}, \mathbf{Y}], \mathbf{Z}]$ and torsion-free. This connection is the average between left and right Cartan connection. Therefore the Cartan

connections of a Lie group are:

$$\begin{aligned}
\tilde{\nabla}_{\tilde{\mathbf{X}}}\tilde{\mathbf{Y}} &= 0 && \text{Left}\\
\tilde{\nabla}_{\tilde{\mathbf{X}}}\tilde{\mathbf{Y}} &= \frac{1}{2}\left[\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right] && \text{Symmetric}\\
\tilde{\nabla}_{\tilde{\mathbf{X}}}\tilde{\mathbf{Y}} &= \left[\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right] && \text{Right}
\end{aligned}$$

2.2 Parallel Transport on Cartan connections

Once described the conditions for the one-parameters subgroups to be geodesics, it is natural to ask how to parallel transport along these paths, and each Cartan connection lead to a specific parallel transport method.

For the left Cartan connection, the unique fields that are covariantly constant are the *left-invariant* vector fields, and the parallel transport is induced by the left multiplication, i.e. $\Pi^L : T_p\mathbb{G} \to T_q\mathbb{G}$ is defined as

$$\Pi^L(\mathbf{X}) = DL_{qp^{-1}}\mathbf{X} \tag{4}$$

Conversely, the *right-invariant* vector fields are covariantly constant with respect to to the right invariant connection. As above, the parallel transport is given by the differential of the right translation $\Pi^R(\mathbf{X}) = DR_{p^{-1}q}\mathbf{X}$. Finally, for the symmetric Cartan connection the parallel transport is given by the combination of the left and right transports. In fact it can be shown [5] that the parallel transport of $\mathbf{X}$ along the curve $\exp(t\mathbf{Y})$ is

$$\Pi^S(\mathbf{X}) = DL_{\exp(\frac{1}{2}\mathbf{Y})}DR_{\exp(\frac{1}{2}\mathbf{Y})}\mathbf{X} \tag{5}$$

3 Application to image registration

The Lie group theory is of relevant interest in the image registration context. For instance, the Lie group exponential has been already used for the diffeomorphic registration parametrized by stationary velocity fields. Of course, when moving to the infinite dimensional group of the diffeomorphisms, some caution is required for the generalization of the standard Lie theory and further research is still needed in order to quantify the impact of using such mathematical framework. However, the effectiveness of the SVF parametrization in terms of registration accuracy and computational efficiency encourage the adoption of the SVF as a valid instrument for the computational anatomy.
Given that one-parameter subgroups are geodesics for all the Cartan connections, we can implement the associated parallel transport for the SVF. Given the left and right actions on $\mathit{diff}(M)$,

$$L_f g = f \circ g \qquad R_f g = g \circ f$$

we have

$$DL_f \simeq Df \cdot g \qquad DR_f \simeq g \circ f$$

We can therefore provide an *explicit closed form formula* for the parallel transport with respect to the canonical Cartan connections. In particular, if $\mathbf{X}$ is a vector to be transported, and $\exp(t\mathbf{Y})$ is the one-parameter subgroup we have:

$$\Pi^L_{\mathbf{Y}}(\mathbf{X}) = D\exp(\mathbf{Y}) \cdot \mathbf{X} \tag{6}$$

$$\Pi^R_{\mathbf{Y}}(\mathbf{X}) = \mathbf{X} \circ \exp(\mathbf{Y}) \tag{7}$$

$$\Pi^S_{\mathbf{Y}}(\mathbf{X}) = D\exp(\frac{\mathbf{Y}}{2}) \cdot \left(\mathbf{X} \circ \exp(\frac{\mathbf{Y}}{2})\right) \tag{8}$$

Remark 1. The geodesics given by the Cartan connection are intrinsically different from the metric Riemannian ones, in the sense that the underlying connection is different from the Levi-Civita one. In particular the geodesics are not related to a positive definite from in $T\mathbb{G} \times T\mathbb{G}$. As consequence the space is not metrically complete, i.e. not all the elements of the space $\mathbb{G}$ might be reached by the one-parameter subgroups. The effect of such geometric property in the image registration context requires further investigation, in order to characterize the transformations that cannot be parametrized by SVF. However, we observe that in the image registration we are not interested in recovering "all" the possible diffeomorphisms, but only those which lead to admissible anatomical transformations.

Remark 2. From the computational point of view, we notice that among the three transport methods, Π^R requires the simple resampling of the velocity field, while both Π^L and Π^S involve the computation of the Jacobian Matrix. The involvement of high order terms can raise accuracy problems, especially in case of noisy data and numerical approximations. We can alleviate the computational inaccuracy by taking advantage of the scaling properties of the one-parameter subgroups. Rather than directly compute the Jacobian $D\exp(\mathbf{Y})$, from the property

$$\begin{aligned}\exp(\mathbf{Y}) &= \exp(\frac{\mathbf{Y}}{2}) \circ \exp(\frac{\mathbf{Y}}{2}) = \\ &= \exp(\frac{\mathbf{Y}}{n}) \circ \ldots \circ \exp(\frac{\mathbf{Y}}{n}).\end{aligned}$$

we can derive an iterative scheme for the Jacobian computation.
In fact, given a suitable first approximation $D\exp(\mathbf{Y})^{[0]} \simeq D\frac{\mathbf{Y}}{n}$ for an opportune scaling factor n, we have the iterative formula

$$D\exp(\mathbf{Y})^{[N+1]} = D\exp(\mathbf{Y})^{[N]}|_{\exp(\frac{\mathbf{Y}}{n})} \cdot D\exp(\frac{\mathbf{Y}}{n}) \tag{9}$$

Thanks to the iterative scheme (9) the Jacobian is updated for a sufficient number of small steps along the one-parameter subgroup. Thus, the scheme avoids the computation of high order quantities on the final deformation field, that could

introduce biases due to the discretization inaccuracy. In fact, the derivatives here are more robustly computed only for an opportunely scaled velocity field, and the iterative formula evaluates the final Jacobian by successive resampling and multiplications. Although the resampling scheme has an important impact on the final computational accuracy, in the following it will be performed by simple scalar interpolation.

4 Transport of longitudinal atrophy

4.1 Synthetic experiment on a simplified geometry

A synthetic progression of longitudinal atrophy was simulated on a simplified geometry, represented by a 3D gray matter sphere S_0 enclosing a white/black matter region. The atrophy was simulated by decreasing the gray layer thickness on four subsequent time points to generate the sequence S_i, $i = 1-4$ (Figure 2). The longitudinal trajectories of deformation fields $\exp(\mathbf{X}^i)$ were then evaluated by registering the images to the baseline with the Log-Demons algorithm [15]. The sequence of deformations $\exp(\mathbf{X}^i)$ was then transported on a target ellipsoidal geometry E_0 along the inter-subject deformation $\exp(\mathbf{Y})$ such that $\exp(\mathbf{Y}) * S_0 = E_0$. The transport methods that we tested were:

- Π^R,
- Π^L and Π^S with the iterative scheme,
- the conjugate action $Conj(\exp(\mathbf{X}^i)) = \exp(\mathbf{Y})\exp(\mathbf{X}^i)\exp(\mathbf{Y})^{-1}$.

Moreover, the velocity fields $\mathbf{X}^i$ were transported with the Schild's Ladder, which operates along the "diagonal" inter-subject deformations $\exp(\mathbf{Y^i})$ such that $\exp(\mathbf{Y^i}) * S_i = E_0$ (Figure 1).

The methods were quantitatively assessed by evaluating the features of interest in the ellipsoid gray layer: the average L^2 Norm of the transported stationary velocity field and the Jacobian determinant, log-Jacobian determinant and Elastic energy of the associated deformation fields. Since we are interested in preserving the interesting features of the transported trajectories, the transported quantities were compared to the original values in the reference sphere space. Moreover, the stability of the methods was tested by checking the scalar spatial maps associated to the features.

Results Table 1 shows the accuracy of the transport methods in the preservation of the measure of changes in the gray matter layer. Among the different methods, the transport Π^R was the most accurate in preserving the average measures, while the Schild's Ladder performed better on the Log-Jacobian.

From the inspection of the related scalar log-Jacobian maps (Figure 2), the transport Π^L is the less stable and leads to noisy maps. Moreover, we notice that the areas of expansions does not fit the boundary of the ellipsoid. On the

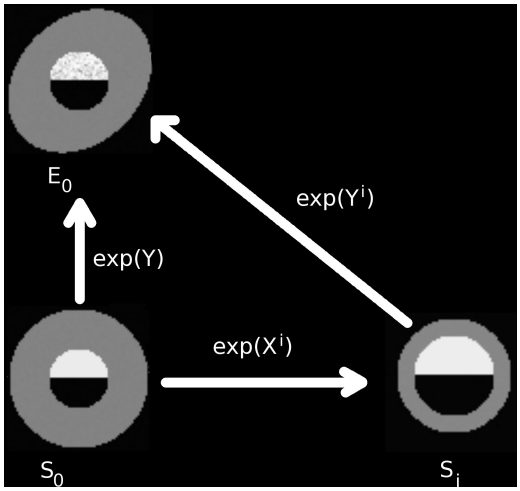


Fig. 1. Synthetic example: Intra and inter-subject variations from the sphere source space to the ellipsoid target space with related deformations.

other hand, the transport Π^R leads to smooth maps of changes, consistent with the target geometry, while the transport Π^S lies "in between", as one could reasonably expect. The Schild's Ladder lead to smooth maps as well, although the inner spherical shape seems corrupted for higher deformations. This could explain the lower performance on the quantitative measurements for the time points 3 and 4. Finally, the log-Jacobian maps associated to the conjugate actions are smooth but fail to preserve the target ellipsoidal geometry, especially for the higher deformations.

Table 1. Average measures of changes on the gray matter layer. Top-row (Source Space): changes measured on the reference sphere at each time point $1-4$. Bottom-rows: changes measured from the transported longitudinal deformations on the ellipsoid. For the conjugate action it was not possible to compute the L^2 Norm of the associated stationary velocity field, since it acts on deformation fields.

	L^2 Norm				Log Jacobian				Jacobian				Elastic energy			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Source Space	2.97	9.85	22.68	44.62	-4.77	-9.54	-14.76	-19.14	0.68	0.47	0.35	0.37	3.47	3.93	4.5	5.23
Π^L	3.02	9.57	22.14	42.32	-5	-9.82	-14.88	-20.43	**0.69**	0.51	0.43	0.45	3.51	4.01	4.67	5.53
Π^R	**2.94**	**10**	**22.81**	**44.58**	-4.70	**-9.36**	-14.51	-19.18	**0.69**	**0.49**	**0.36**	**0.37**	**3.49**	**3.9**	**4.44**	**5.15**
Π^S	3.3	11.17	25.7	50.37	-5.74	-11.2	-17.13	-23.65	**0.67**	0.50	0.42	0.48	3.58	4.2	4.99	6.05
Schild's Ladder	3.65	10.74	24.3	51.49	**-4.83**	-9.86	**-14.65**	**-19.11**	0.71	0.51	0.45	0.49	3.57	4.14	4.84	6.21
Conjugate	/	/	/	/	-2.6	-5.5	-9.18	-13.93	0.8	0.63	0.47	0.32	3.43	3.83	4.36	5.04

5 Conclusions

The study shows how the straightforward application of the Lie group theory to the diffeomorphic registration can lead to simple and efficient solutions for the transport of deformations. In particular the one-parameters subgroups are the geodesics with respect to the Cartan connections, and this mathematical setting leads to a closed form solution for the parallel transport. The geodesic of the Cartan connections generally differ from those of the Riemannian framework like the LDDMM, in the sense that they are not defined from a metric on the tangent space, and consequently the parallel transport is not related to the preservation of metric properties.
The present study highlights the trade-off between the choice of proper mathematical constructions and the related numerical implementation. In fact, among the parallel transports from the Cartan connections, the right one showed greater accuracy and smoothness, due to the simple computational requirements. However, the transport Π^R operates according to a specific geometry corresponding to the right Cartan connection. In this case, we are working in a zero-curvature space with torsion, while from the theoretical point of view it might be preferable to work with respect to a symmetric connection which leads to torsion-free spaces. At this purpose, further studies are required in order to clarify the effects in the image registration context of imposing a specific connection, and for defining more robust numerical schemes for the computation of high order quantities.

Acknowledgements This work was partially supported by the ANR program *Karametria* ANR-09-BLAN-0332 from the French Agence Nationale pour la Recherche.

References

1. V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-euclidean framework for statistics on diffeomorphisms. In R. Larsen, M. Nielsen, and J. Sporring, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 4190, page 924. Springer, Heidelberg, Sep 2006.
2. M.N. Bossa, E. Zacur, and S. Olmos. On changing coordinate systems for longitudinal tensor-based morphometry. *Spatio Temporal Image Analysis Workshop (STIA), MICCAI 2010*, 2010.
3. S. Durrleman, P. Fillard, X. Pennec, A. Trouve, and N. Ayache. Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage*, 2011.
4. S. Durrleman, X. Pennec, A. Trouv, G. Gerig, and N. Ayache. Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In G. Yang, D. Hawkes, D. Rueckert, J Noble, and C. Taylor, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 5761. Springer, Heidelberg, Sep 2009.
5. S. Helgason. *Differential geometry, Lie groups, and symmetric spaces*. American Mathematical Soc, 1978.
6. M. Lorenzi, N. Ayache, and X. Pennec. Schilds ladder for the parallel transport of deformations in time series of images. In G. Székely and H. Hahn, editors, *Information Processing in Medical Imaging - IPMI*. Springer, Heidelberg, Jul 2011.

7. M. Lorenzi, G.B Frisoni, N. Ayache, and X. Pennec. Mapping the effects of aβ_{1-42} levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields. In G. Fichtinger, A. Martel, and T. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. Springer, Heidelberg, Sep 2011.
8. T. Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache. Logdemons revisited: Consistent regularisation and incompressibility constraint for soft tissue tracking in medical images. In T. Jiang, N. Navab, J.P.W. Pluim, and M. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. Springer, Heidelberg, Sep 2010.
9. T. Mansi, I. Voigt, B. Leonardi, X. Pennec, S. Durrleman, M. Sermesant, H. Delingette, A.M. Taylor, Y. Boudjemline, G. Pongiglione, and N. Ayache. A statistical model for quantification and prediction of cardiac remodelling: Application to tetralogy of fallot. *IEEE Transactions on Medical Images*, 2011.
10. M. Miller, A. Trouvé, and L. Younes. On the metrics and Euler-Lagrange equations of computational anatomy. *Annu Rev Biomed Eng*, 4(1):375–405, 2002.
11. M.M. Postnikov. *Geometry VI*. Springer, 2001.
12. A. Rao, R. Chandrashekara, G. Sanchez-Hortiz, R. Mohiaddin, P. aljabar, J. Hajnal, B. Puri, and D. Rueckert. Spatial trasformation of motion and deformation fields using nonrigid registration. *IEEE Transactions on Medical Imaging*, 23(9), 2004.
13. C. Seiler, X. Pennec, and M. Reyes. Geometry-Aware Multiscale Image Registration Via OBBTree-Based Polyaffine Log-Demons. In G. Fichtinger, A. Martel, and T. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. Springer, Heidelberg, Sep 2011.
14. P. Thompson, K.M. Ayashi, G. Zubicaray, A.L. Janke, S.E. Rose, J. Semple, D. Herman, M.S. Hong, S.S. Dittmer, D.M. Dodrell, and A.W. Toga. Dynamics of gray matter loss in alzheimer's disease. *The Journal of Neuroscience*, 23:994–1005, 2003.
15. T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Symmetric Log-domain diffeomorphic registration: A Demons-based approach. In D. Metaxas, L. Axel, G. Fichtinger, and G. Szekely, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 5241, pages 754–761. Springer, Heidelberg, Sep 2008.
16. L. Younes, A. Qiu, R. Winslow, and M. Miller. Transport of relational structures in groups of diffeomorphisms. *J Math Imaging Vis*, 32(1):41–56, Sep 2008.

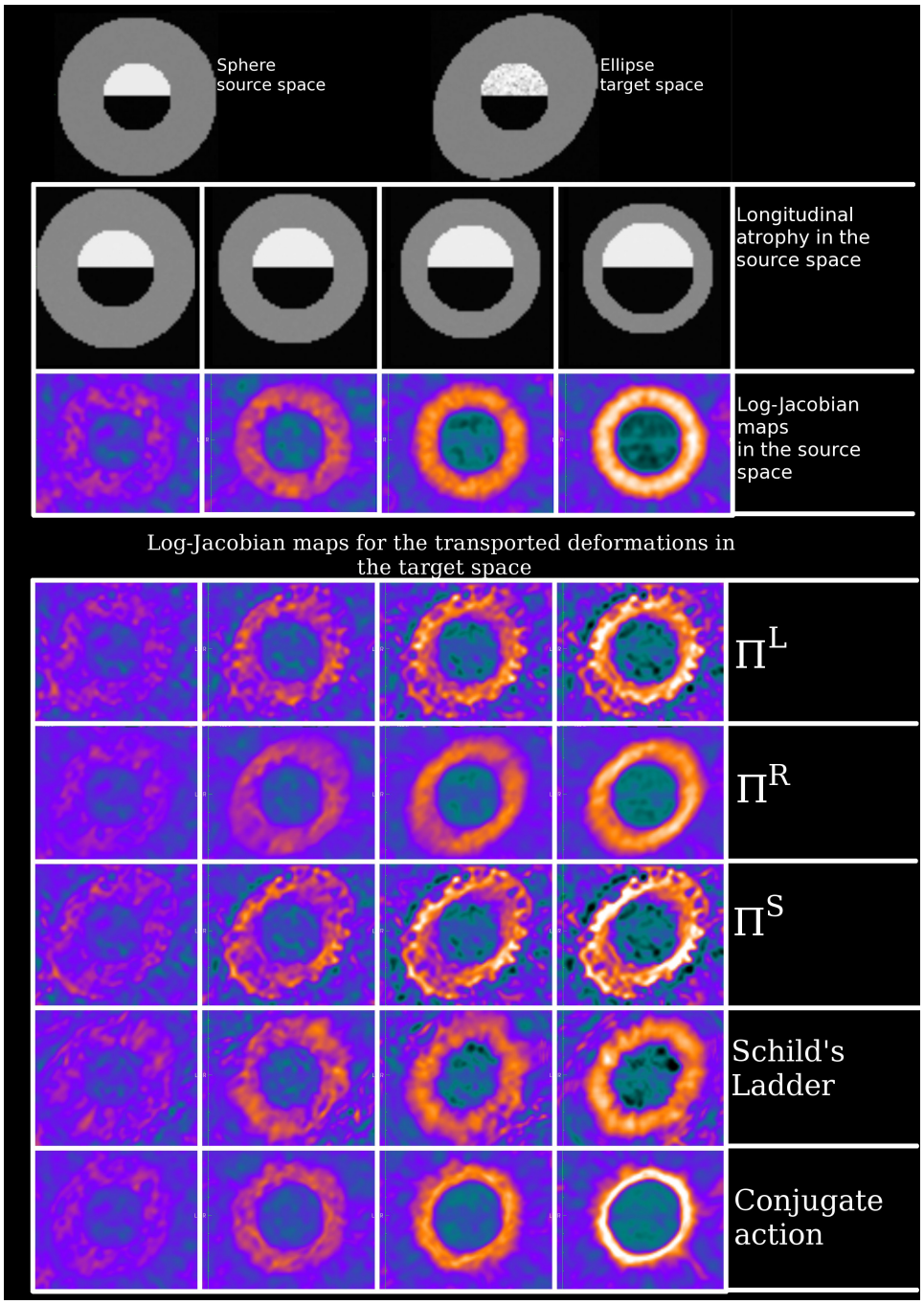


Fig. 2. Top row: Spherical source and ellipsoidal target geometrical references. From top to bottom: Longitudinal atrophy sequence in the spherical space, associated log-Jacobian determinant scalar maps, and log-Jacobian determinant maps associated to the different methods of transport.

Geodesic Regression on Riemannian Manifolds

P. Thomas Fletcher

School of Computing, University of Utah

Abstract. This paper introduces a regression method for modeling the relationship between a manifold-valued random variable and a real-valued independent parameter. The principle is to fit a geodesic curve, parameterized by the independent parameter, that best fits the data. Error in the model is evaluated as the sum-of-squared geodesic distances from the model to the data, and this provides an intrinsic least squares criterion. Geodesic regression is, in some sense, the simplest parametric model that one could choose, and it provides a direct generalization of linear regression to the manifold setting. A hypothesis test for determining the significance of the estimated trend is also developed. While the method can be generally applied to data on any manifold, specific examples are given for a set of synthetically generated rotation data and an application to analyzing shape changes in the corpus callosum due to age.

1 Introduction

Regression analysis is a fundamental statistical tool for determining how a measured variable is related to one or more potential explanatory variables. The most widely used regression model is linear regression, due to its simplicity, ease of interpretation, and ability to model many phenomena. However, if the response variable takes values on a nonlinear manifold, a linear model is not applicable. Such manifold-valued measurements arise in many applications, including those involving directional data, transformations, tensors, and shape. For example, in biology and medicine it is often critical to understand processes that change the shape of anatomy. The difficulty is that shape variability is inherently high-dimensional and nonlinear. An effective approach to capturing this variability has been to parameterize shape as a manifold, or shape space.

Several works have studied the regression problem on manifolds. Jupp and Kent [6] propose an unrolling method on shape spaces. Regression analysis on the group of diffeomorphisms has been proposed as growth models by Miller [10], nonparametric regression by Davis, et al. [3], and second order splines by Trouvé and Vialard [12]. Finally, Shi, et al. [11] proposed a semiparametric model with multiple covariates for manifold response data. None of these methods provide a direct generalization of linear regression to manifolds. The purpose of this paper is to develop such a generalization, called geodesic regression, which models the relationship between an independent scalar variable with a dependent manifold-valued random variable as a geodesic curve. Like linear regression, the advantages of this model are its simplicity and ease of interpretation. As will be shown, the geodesic regression model also leads to a straightforward generalization of the R^2 statistic and a hypothesis test for significance of the estimated geodesic trend.

2 Multiple Linear Regression

Before formulating geodesic regression on general manifolds, we begin by reviewing multiple linear regression in $\mathbb{R}^n$. Here we are interested in the relationship between a non-random *independent* variable $X \in \mathbb{R}$ and a random *dependent* variable Y taking values in $\mathbb{R}^n$. A multiple linear model of this relationship is given by

$$Y = \alpha + X\beta + \epsilon, \tag{1}$$

where $\alpha \in \mathbb{R}^n$ is an unobservable *intercept* parameter, $\beta \in \mathbb{R}^n$ is an unobservable *slope* parameter, and ϵ is an $\mathbb{R}^n$-valued, unobservable random variable representing the error. Geometrically, this is the equation of a one-dimensional line through $\mathbb{R}^n$ (plus noise), parameterized by the scalar variable X. For the purposes of generalizing to the manifold case, it is useful to think of α as the starting point of the line and β as a velocity vector.

Given realizations of the above model, i.e., data $(x_i, y_i) \in \mathbb{R} \times \mathbb{R}^n$, for $i = 1, \ldots, N$, the least squares estimates, $\hat{\alpha}, \hat{\beta}$, for the intercept and slope are computed by solving the minimization problem

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{(\alpha,\beta)} \sum_{i=1}^{N} \|y_i - \alpha - x_i\beta\|^2 . \tag{2}$$

This equation can be solved analytically, yielding

$$\hat{\beta} = \frac{\frac{1}{N}\sum x_i\, y_i - \bar{x}\,\bar{y}}{\sum x_i^2 - \bar{x}^2},$$
$$\hat{\alpha} = \bar{y} - \bar{x}\,\hat{\beta},$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the x_i and y_i, respectively. If the errors in the model are drawn from distributions with zero mean and finite variance, then these estimators are unbiased and consistent.

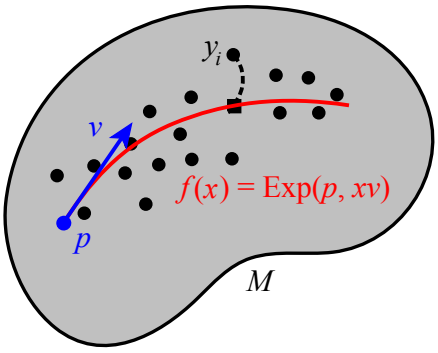


Fig. 1. Schematic of the geodesic regression model.

3 Geodesic Regression

Let $y_1, \ldots, y_N$ be points on a smooth Riemannian manifold M, with associated scalar values $x_1, \ldots, x_N \in \mathbb{R}$. The goal of geodesic regression is to find a geodesic curve γ on M that best models the relationship between the x_i and the y_i. Just as in linear regression, the speed of the geodesic will be proportional to the independent parameter corresponding to the x_i. Estimation will be set up as a least-squares problem, where we want to minimize the sum-of-squared Riemannian distances between the model and the data. A schematic of the geodesic regression model is shown in Figure 1.

Before formulating the model, we review a few basic concepts of Riemannian geometry. We will write an element of the tangent bundle as the pair $(p, v) \in TM$, where p is a point in M and $v \in T_pM$ is a tangent vector at p. Recall that for any $(p, v) \in TM$ there is a unique geodesic curve γ, with initial conditions $\gamma(0) = p$ and $\gamma'(0) = v$. This geodesic is only guaranteed to exist locally. When γ is defined over the interval $[0, 1]$, the exponential map at p is defined as $\mathrm{Exp}_p(v) = \gamma(1)$. In other words, the exponential map takes a position and velocity as input and returns the point at time 1 along the geodesic with these initial conditions. The exponential map is locally diffeomorphic onto a neighborhood of p. Let $V(p)$ be the largest such neighborhood. Then within $V(p)$ the exponential map has an inverse, the Riemannian log map, $\mathrm{Log}_p : V(p) \to T_pM$. For any point $q \in V(p)$ the Riemannian distance function is given by $d(p, q) = \| \mathrm{Log}_p(q) \|$. It will be convenient to include the point p as a parameter in the exponential and log maps, i.e., define $\mathrm{Exp}(p, v) = \mathrm{Exp}_p(v)$ and $\mathrm{Log}(p, q) = \mathrm{Log}_p(q)$.

Notice that the tangent bundle TM serves as a convenient parameterization of the set of possible geodesics on M. An element $(p, v) \in TM$ provides an intercept p and a slope v, analogous to the α and β parameters in the multiple linear regression model (1). In fact, β is a vector in the tangent space $T_\alpha \mathbb{R}^n \cong \mathbb{R}^n$, and thus (α, β) is an element of the tangent bundle $T\mathbb{R}^n$. Now consider an M-valued random variable Y and a non-random variable $X \in \mathbb{R}$. The generalization of the multiple linear model to the manifold setting is the *geodesic model*,

$$Y = \mathrm{Exp}(\mathrm{Exp}(p, Xv), \epsilon), \tag{3}$$

where ϵ is a random variable taking values in the tangent space at $\mathrm{Exp}(p, Xv)$. Notice that for Euclidean space, the exponential map is simply addition, i.e., $\mathrm{Exp}(p, v) = p + v$. Thus, the geodesic model coincides with (1) when $M = \mathbb{R}^n$.

3.1 Least Squares Estimation

Consider a realization of the model (3): $(x_i, y_i) \in \mathbb{R} \times M$, for $i = 1, \ldots, N$. Given this data, we wish to find estimates of the parameters $(p, v) \in TM$. First, define the sum-of-squared error of the data from the geodesic given by (p, v) as

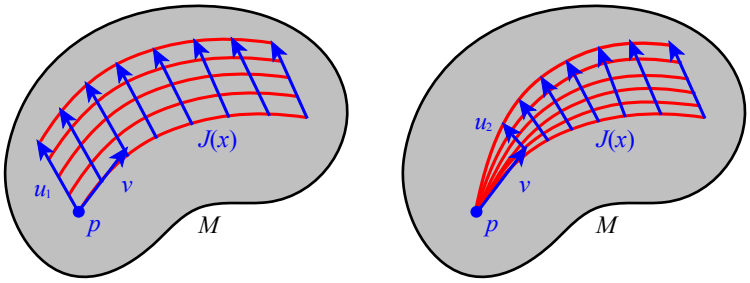
$$E(p, v) = \frac{1}{2} \sum_{i=1}^{N} d(\mathrm{Exp}(p, x_i v), y_i)^2. \tag{4}$$

Following the ordinary least squares minimization problem given by (2), we formulate a least squares estimator of the geodesic model as a minimizer of the above sum-of-squares energy, i.e.,

$$(\hat{p}, \hat{v}) = \arg\min_{(p,v)} E(p, v). \tag{5}$$

Again, notice that this problem coincides with the ordinary least squares problem when $M = \mathbb{R}^n$.

Unlike the linear setting, the least squares problem in (5) for a general manifold M will typically not yield an analytic solution. Instead we derive a gradient descent algorithm. Computation of the gradient of (4) will require two parts: the derivative of the Riemannian distance function and the derivative of the exponential map. Fixing a point $p \in M$, the gradient of the squared distance function is $\nabla_x d(p, x)^2 = -2 \operatorname{Log}_x(p)$ for $x \in V(p)$.



$d_p \operatorname{Exp}$ $d_v \operatorname{Exp}$

Fig. 2. Jacobi fields as derivatives of the exponential map.

The derivative of the exponential map $\operatorname{Exp}(p, v)$ can be separated into a derivative with respect to the initial point p and a derivative with respect to the initial velocity v. To do this, first consider a variation of geodesics given by $c_1(s, t) = \operatorname{Exp}(\operatorname{Exp}(p, su_1), tv(s))$, where $u_1 \in T_pM$ defines a variation of the initial point along the geodesic $\eta(s) = \operatorname{Exp}(p, su_1)$. Here we have also extended $v \in T_pM$ to a vector field $v(s)$ along η via parallel translation. This variation is illustrated on the left side of Figure 2. Next consider a variation of geodesics $c_2(s, t) = \operatorname{Exp}(p, su_2 + tv)$, where $u_2 \in T_pM$. (Technically, u_2 is a tangent to the tangent space, i.e., an element of $T_v(T_pM)$, but there is a natural isomorphism $T_v(T_pM) \cong T_pM$.) The variation c_2 produces a "fan" of geodesics as seen on the right side of Figure 2.

Now the derivatives of $\operatorname{Exp}(p, v)$ with respect to p and v are given by

$$d_p \operatorname{Exp}(p, v) \cdot u_1 = \frac{d}{ds} c_1(s, t)\Big|_{s=0} = J_1(1)$$
$$d_v \operatorname{Exp}(p, v) \cdot u_2 = \frac{d}{ds} c_2(s, t)\Big|_{s=0} = J_2(1),$$

where $J_i(t)$ are *Jacobi fields* along the geodesic $\gamma(t) = \mathrm{Exp}(p, tv)$. Jacobi fields are solutions to the second order equation

$$\frac{D^2}{dt^2} J(t) + R(J(t), \gamma'(t))\ \gamma'(t) = 0, \tag{6}$$

where R is the Riemannian curvature tensor. For more details on the derivation of the Jacobi field equation and the curvature tensor, see for instance [2]. The initial conditions for the two Jacobi fields above are $J_1(0) = u_1, J_1'(0) = 0$ and $J_2(0) = 0, J_2'(0) = u_2$, respectively. If we decompose the Jacobi field into a component tangential to γ and a component orthogonal, i.e., $J = J^\top + J^\perp$, the tangential component is linear: $J^\top(t) = u_1^\top + t u_2^\top$. Therefore, the only challenge is to solve for the orthogonal component.

Finally, the gradient of the sum-of-squares energy in (4) is given by

$$\nabla_p E(p, v) = -\sum_{i=1}^{N} d_p \,\mathrm{Exp}(p, x_i v)^\dagger \,\mathrm{Log}(\mathrm{Exp}(p, x_i v), y_i),$$
$$\nabla_v E(p, v) = -\sum_{i=1}^{N} x_i\, d_v \,\mathrm{Exp}(p, x_i v)^\dagger \,\mathrm{Log}(\mathrm{Exp}(p, x_i v), y_i),$$

where we have taken the adjoint of the exponential map derivative, e.g., defined by $\langle d_p \,\mathrm{Exp}(p, v)u, w\rangle = \langle u, d_p \,\mathrm{Exp}(p, v)^\dagger w\rangle$. As we will see in the next section, formulas for Jacobi fields and their respective adjoint operators can often be derived analytically for many useful manifolds.

3.2 R^2 Statistics and Hypothesis Testing

In regression analysis the most basic question one would like to answer is whether the relationship between the independent and dependent variables is significant. A common way to test this is to see if the amount of variance explained by the model is high. For geodesic regression we will measure the amount of explained variance using a generalization of the R^2 statistic, or coefficient of determination, to the manifold setting. To do this, we first define predicted values of y_i and the errors ϵ_i as

$$\hat{y}_i = \mathrm{Exp}(\hat{p}, x_i \hat{v}),$$
$$\hat{\epsilon}_i = \mathrm{Log}(\hat{y}_i, y_i),$$

where $(\hat{p}, \hat{v})$ are the least squares estimates of the geodesic parameters defined above. Note that the $\hat{y}_i$ are points along the estimated geodesic that are the best predictions of the y_i given only the x_i. The $\hat{\epsilon}_i$ are the residuals from the model predictions to the true data.

Now to define the total variance of data, $y_1, \ldots, y_N \in M$, we use the Fréchet variance, intrinsically defined by

$$\mathrm{var}(y_i) = \min_{y \in M} \frac{1}{N} \sum_{i=1}^{N} d(y, y_i)^2.$$

The unexplained variance is the variance of the residuals, $\text{var}(\hat{\epsilon}_i) = \frac{1}{N}\sum \|\hat{\epsilon}_i\|^2$. From the definition of the residuals, it can be seen that the unexplained variance is the mean squared distance of the data to the model, i.e., $\text{var}(\hat{\epsilon}_i) = \frac{1}{N}\sum d(\hat{y}_i, y_i)^2$. Using these two variance definitions, the generalization of the R^2 statistic is then given by

$$R^2 = 1 - \frac{\text{unexplained variance}}{\text{total variance}} = 1 - \frac{\text{var}(\hat{\epsilon}_i)}{\text{var}(y_i)}. \tag{7}$$

Fréchet variance coincides with the standard definition of variance when $M = \mathbb{R}^n$. Therefore, it follows that the definition of R^2 in (7) coincides with the R^2 for linear regression when $M = \mathbb{R}^n$. Also, because Fréchet variance is always nonnegative, we see that $R^2 \leq 1$, and that $R^2 = 1$ if and only if the residuals to the model are exactly zero, i.e., the model perfectly fits the data. Finally, it is clear that the residual variance is always smaller than the total variance, i.e., $\text{var}(\hat{\epsilon}_i) \leq \text{var}(y_i)$. This is because we could always choose $\hat{p}$ to be the Fréchet mean and $v = 0$ to achieve $\text{var}(\hat{\epsilon}_i) = \text{var}(y_i)$. Therefore, $R^2 \geq 0$, and it must lie in the interval $[0, 1]$, as is the case for linear models.

We now describe a permutation test for testing the significance of the estimated slope term, $\hat{v}$. Notice that if we constrain v to be zero in (5), then the resulting least squares estimate of the intercept, $\hat{p}$, will be the Fréchet mean of the y_i. The desired hypothesis test is whether the fraction of unexplained variance is significantly decreased by also estimating v. The null hypothesis is $H_0 : R^2 = 0$, which is the case if the unexplained variance in the geodesic model is equal to the total variance. Under the null hypothesis, there is no relationship between the X variable and the Y variable. Therefore, the x_i are exchangeable under the null hypothesis, and a permutation test may randomly reorder the x_i data, keeping the y_i fixed. Estimating the geodesic regression parameters for each random permutation of the x_i, we can calculate a sequence of R^2 values, $R^2_1, \ldots, R^2_m$, which approximate the sampling distribution of the R^2 statistic under the null hypothesis. Computing the fraction of the R^2_k that are greater than the R^2 estimated from the unpermuted data gives us a p-value.

4 Results

4.1 Regression of 3D Rotations

Overview of Unit Quaternions We represent 3D rotations as the unit quaternions, $\mathbb{Q}_1$. A quaternion is denoted as $q = (a, v)$, where a is the "real" component and $v = bi + cj + dk$. Geodesics in the rotation group are given simply by constant speed rotations about a fixed axis. Let $e = (1, 0)$ be the identity quaternion. The tangent space $T_e\mathbb{Q}_1$ is the vector space of quaternions of the form $(0, v)$. The tangent space at an arbitrary point $q \in \mathbb{Q}_1$ is given by right multiplication of $T_e\mathbb{Q}_1$ by q. The Riemannian exponential map is $\text{Exp}_q((0, v) \cdot q) = (\cos(\theta/2), 2v \cdot \sin(\theta/2)/\theta) \cdot q$, where $\theta = 2\|v\|$. The log map is given by $\text{Log}_q((a, v) \cdot q) = (0, \theta v/\|v\|) \cdot q$, where $\theta = \arccos(a)$.

Being a unit sphere, $\mathbb{Q}_1$ has constant sectional curvature $K = 1$. In this case the orthogonal component of the Jacobi field equation (6) along a geodesic $\gamma(t)$

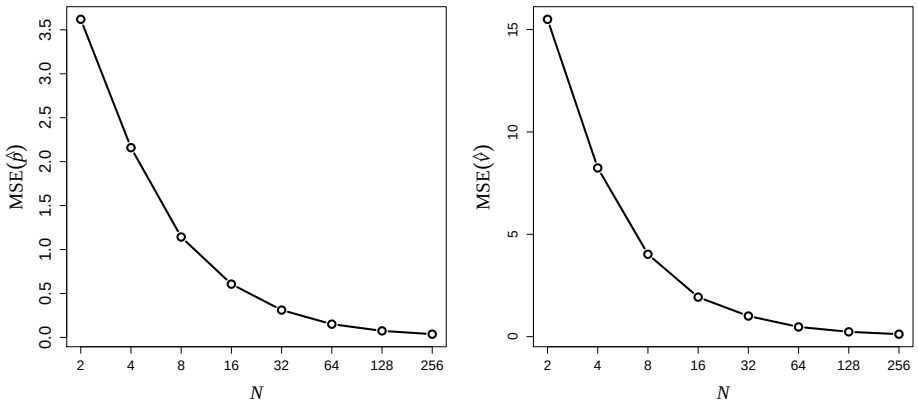


Fig. 3. Results for simulated rotation data: MSE of the geodesic regression estimates for the intercept (left) and slope (right) as a function of sample size.

has the analytic solution

$$J(t)^{\perp} = u_1(t)\cos(Lt) + u_2(t)\sin(Lt)\,,$$

where u_1, u_2 are parallel vector fields along γ, with initial conditions $u_1(0) = J(0)^{\perp}$ and $u_2(0) = J'(0)^{\perp}$, and $L = \|\gamma'\|$. While the Jacobi field equation gives us the differential of the exponential map, we really need the adjoint of this operator for geodesic regression. However, from the above equation it is clear that $d_p \operatorname{Exp}$ and $d_v \operatorname{Exp}$ are both self-adjoint operators. That is, the above Jacobi field equation provides us both the differential and its adjoint.

Geodesic Regression of Simulated Rotation Data To test the geodesic regression least squares estimation on $\mathbb{Q}_1$, synthetic rotation data was simulated according to the geodesic model (3). The intercept was the identity rotation: $p = (1, 0, 0, 0)$, and the slope was a rotation about the z-axis: $v = (0, 0, 0, \pi/4)$. The x_i data were drawn from a uniform distribution on $[0, 1]$. The errors in the model were generated from an isotropic Gaussian distribution in the tangent space, with $\sigma = \pi/8$. The resulting data (x_i, y_i) were used to compute estimates of the parameters $(\hat{p}, \hat{v})$. This experiment was repeated 1,000 times each for sample sizes $N = 2^k$, $k = 1, \ldots, 8$. We would expect that as the sample size increases, the mean squared error (MSE) in the estimates $(\hat{p}, \hat{v})$, relative to the true parameters, would approach zero. The MSE is defined as

$$\operatorname{MSE}(\hat{p}) = \frac{1}{M}\sum_{i=1}^{M} d(\hat{p}_i, p)^2, \qquad \operatorname{MSE}(\hat{v}) = \frac{1}{M}\sum_{i=1}^{M} \|\hat{v}_i \cdot (\hat{p}_i^{-1} p) - v\|^2,$$

where $M = 1{,}000$ is the number of repeated trials, and $(\hat{p}_i, \hat{v}_i)$ is the estimate from the ith trial. Notice the multiplication by $(\hat{p}_i^{-1} p)$ in the second equation

is a right-translation of $\hat{v}_i$ to the tangent space of p. Figure 3 shows plots of the resulting MSE for the slope and intercept estimates. As expected, the MSE approaches zero as sample size increases, indicating at least empirically that the least squares estimates are consistent.

4.2 Regression in Shape Spaces

One area of medical image analysis and computer vision that finds the most widespread use of Riemannian geometry is the analysis of shape. Dating back to the groundbreaking work of Kendall [7] and Bookstein [1], modern shape analysis is concerned with the geometry of objects that is invariant to rotation, translation, and scale. This typically results in representing an object's shape as a point in a nonlinear Riemannian manifold, or *shape space*. Recently, there has been a great amount of interest in Riemannian shape analysis, and several shape spaces for 2D and 3D objects have been proposed [5, 8, 9, 13]. We choose here to use Kendall's shape space, but geodesic regression is applicable to other shape spaces as well. It could also be applied to spaces of diffeomorphisms, using the Jacobi field calculations given by Younes [14].

Overview of Kendall's Shape Space We begin with derivations of the necessary computations for geodesic regression on Kendall's shape space. A configuration of k points in the 2D plane is considered as a complex k-vector, $z \in \mathbb{C}^k$. Removing translation, by requiring the centroid to be zero, projects this point to the linear complex subspace $V = \{z \in \mathbb{C}^k : \sum z_i = 0\}$, which is equivalent to the space $\mathbb{C}^{k-1}$. Next, points in this subspace are deemed equivalent if they are a rotation and scaling of each other, which can be represented as multiplication by a complex number, $\rho e^{i\theta}$, where ρ is the scaling factor and θ is the rotation angle. The set of such equivalence classes forms the complex projective space, $\mathbb{C}P^{k-2}$. As Kendall points out, there is no unique way to identify a shape with a specific point in complex projective space. However, if we consider that the geodesic regression problem only requires computation of exponential/log maps and Jacobi fields, we can formulate these computations without making an explicit identification of shapes with points in $\mathbb{C}P^{k-2}$.

Thus, we think of a centered shape $x \in V$ as representing the complex line $L_x = \{z \cdot x : z \in \mathbb{C}\backslash\{0\}\}$, i.e., L_x consists of all point configurations with the same shape as x. A tangent vector at $L_x \in V$ is a complex vector, $v \in V$, such that $\langle x, v\rangle = 0$. The exponential map is given by rotating (within V) the complex line L_x by the initial velocity v, that is,

$$\mathrm{Exp}_x(v) = \cos\theta \cdot x + \frac{\|x\| \sin\theta}{\theta} \cdot v, \quad \theta = \|v\|. \tag{8}$$

Likewise, the log map between two shapes $x, y \in V$ is given by finding the initial velocity of the rotation between the two complex lines L_x and L_y. Let $\pi_x(y) = x \cdot \langle x, y\rangle / \|x\|^2$ denote the projection of the vector y onto x. Then the log map is given by

$$\mathrm{Log}_x(y) = \frac{\theta \cdot (y - \pi_x(y))}{\|y - \pi_x(y)\|}, \quad \theta = \arccos \frac{|\langle x, y\rangle|}{\|x\|\|y\|}. \tag{9}$$

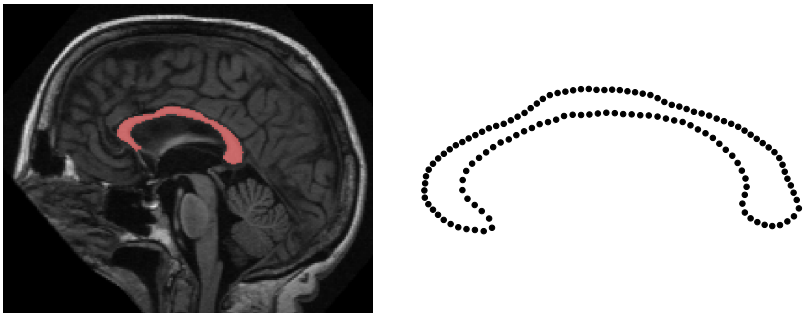


Fig. 4. Corpus callosum segmentation and boundary point model for one subject.

Notice that we never explicitly project a shape onto $\mathbb{C}P^{k-2}$. This has the effect that shapes computed via the exponential map (8) will have the same orientation and scale as the base point x. Also, tangent vectors computed via the log map (9) are valid only at the particular representation x (and not at a rotated or scaled version of x). This works nicely for our purposes and implies that shapes along the estimated geodesic will have the same orientation and scale as the intercept shape, $\hat{p}$.

The sectional curvature of $\mathbb{C}P^{k-2}$ can be computed as follows. Let u, w be orthonormal vectors at a point $p \in \mathbb{C}P^{k-2}$. These vectors may be thought of as vectors in $\mathbb{C}^{k-1} \cong \mathbb{R}^{2k-2}$. Writing the vector w as $w = (w_1, \ldots, w_{2k-2})$, define the operator

$$j(w) = (-w_k, \ldots, -w_{2k-2}, w_1, \ldots, w_{k-1}).$$

(This is just multiplication by $i = \sqrt{-1}$ if we take w as a complex vector with the $k-1$ real coordinates listed first.) Using this operator, the sectional curvature is given by

$$K(u, w) = 1 + 3\langle u, j(w)\rangle^2.$$

When $k = 3$, $\mathbb{C}P^1$ is the space of triangle shapes and is isomorphic to the sphere, S^2, and thus has constant sectional curvature, $K = 1$. For $k > 3$, $\mathbb{C}P^{k-2}$ has sectional curvature in the interval $K \in [1, 4]$. Furthermore, let $u \in T_p\mathbb{C}P^{k-2}$ be any unit length vector. If we decompose the tangent space into an orthonormal basis $e_1, \ldots, e_{2k-2}$, such that $e_1 = j(u)$, then we have $K(u, e_1) = 4$ and $K(u, e_i) = 1$ for $i > 1$. This leads to the following procedure for computing the Jacobi field equation on $\mathbb{C}P^{k-2}$ along a geodesic γ. Given initial conditions for $J(0)^\perp$ and $J'(0)^\perp$, decompose $J(0)^\perp = u_1 + w_1$, so that u_1 is orthogonal to $j(\gamma')$ and w_1 is tangential to $j(\gamma')$. Do the same for $J'(0)^\perp = u_2 + w_2$. As before, extend these vectors to parallel fields, $u_i(t), w_i(t)$, along γ. Then the orthogonal component of the Jacobi field along γ is given by

$$J(t)^\perp = u_1(t)\cos(Lt) + u_2(t)\sin(Lt) + w_1(t)\frac{\cos(2Lt)}{2} + w_2(t)\frac{\sin(2Lt)}{2}.$$

As was the case for rotations, both $d_p \operatorname{Exp}$ and $d_v \operatorname{Exp}$ are self-adjoint operators.

Application to Corpus Callosum Aging The corpus callosum is the major white matter bundle connecting the two hemispheres of the brain. A midsagittal

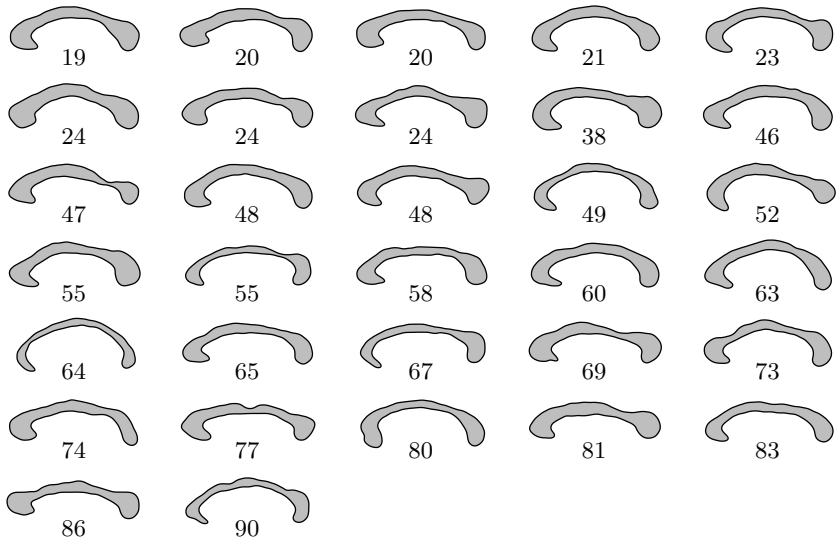


Fig. 5. The input corpus callosum shape data and corresponding subject ages in years.

slice from a magnetic resonance image (MRI) with segmented corpus callosum is shown in Figure 4. Several studies have shown that the volume of the corpus callosum decreases with normal aging [4]. However, less is known about how the *shape* of the corpus callosum changes with age. Understanding shape changes may provide a deeper understanding of the anatomical and biological processes underlying aging. For example, does the corpus callosum shrink uniformly in size, or do certain regions deteriorate faster than others? This type of question can be answered by geodesic regression in shape spaces.

To understand age-related changes in the shape of the corpus callosum, geodesic regression was applied to corpus callosum shape data derived from the OASIS brain database (`www.oasis-brains.org`). The data consisted of MRI from 32 subjects with ages ranging from 19-90 years old. The corpus callosum was segmented in a midsagittal slice using the ITK SNAP program (`www.itksnap.org`). These boundaries of these segmentations were sampled with 128 points using ShapeWorks (`www.sci.utah.edu/software.html`). This algorithm generates a sampling of a set of shape boundaries while enforcing correspondences between different point models within the population. An example of a segmented corpus callosum and the resulting boundary point model is shown in Figure 4. The entire collection of input shapes and their ages is shown in Figure 5 (boundary points have been connected into a boundary curve for visualization purposes). Each of these preprocessing steps were done without consideration of the subject age, to avoid any bias in the data generation.

Geodesic regression was applied to the data (x_i, y_i), where x_i was the ith subject's age, and y_i was the ith subject's corpus callosum, generated as above and represented as a point in Kendall's shape space. First, the average age of the group, $\bar{x}$, was subtracted from each x_i, which was done to make the intercept

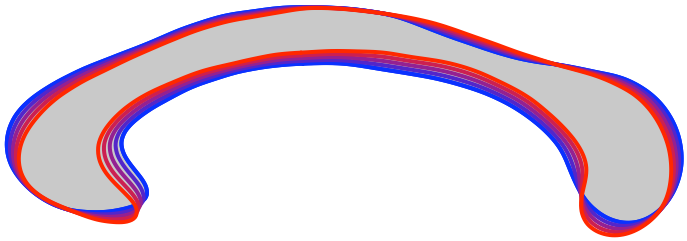


Fig. 6. Geodesic regression of the corpus callosum. The estimated geodesic is shown as a sequence of shapes from age 19 (blue) to age 90 (red).

term correspond to the shape at the mean age, rather than the shape at age zero, which would be far outside the data range. Least squares estimates $(\hat{p}, \hat{v})$ were generated according to (5), and using the above calculations for $\mathbb{C}P^{k-2}$. The resulting estimated geodesic is shown in Figure 6 as a sequence of shapes: $\hat{\gamma}(t_k) = \text{Exp}(\hat{p}, (t_k - \bar{x})\hat{v})$, for $t_k = 19, 36, 54, 72, 90$. The shape trend shows a very clear thinning of the corpus callosum, with the largest effects in the posterior part of the body and in the genu (anterior end).

Finally, the statistical significance of the estimated trend was tested using the permutation test described in Section 3.2, using 10,000 permutations. The p-value for the significance of the slope estimate, $\hat{v}$, was $p = 0.009$. The coefficient of determination (for the unpermuted data) was $R^2 = 0.12$. The low R^2 value must be interpreted carefully. It says that age only describes a small fraction of the shape variability in the corpus callosum. This is not surprising: we would expect the intersubject variability in corpus callosum shape to be difficult to fully describe with a single variable (age). However, this does not mean that the age effects are not important. In fact, the low p-value says that the estimated age changes are highly unlikely to have been found by random chance.

5 Conclusion

We introduced a geodesic regression analysis method for Riemannian manifolds. The geodesic regression model is the natural generalization of linear regression and is parameterized by an intercept and slope term. We also developed a generalization of the R^2 statistic and a permutation test for the significance of the estimated geodesic trend. There are several avenues for future work. First, the hypothesis test presented here could be extended to test for group differences, for example, to test if age-related anatomical changes are different in a disease population compared to controls. Second, theoretical properties of geodesic regression, such as unbiasedness and consistency, would be of interest. Finally, regression diagnostics and model selection procedures need to be developed to assess the appropriateness of a geodesic model for a particular data set.

Acknowledgments: This work was supported by NSF CAREER Grant 1054057.

References

1. Bookstein, F.L.: Size and shape spaces for landmark data in two dimensions (with discussion). Statistical Science 1(2), 181–242 (1986)
2. do Carmo, M.: Riemannian geometry. Birkhäuser (1992)
3. Davis, B., Fletcher, P.T., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: Proceedings of IEEE International Conference on Computer Vision (2007)
4. Driesen, N., Raz, N.: The influence of sex, age, and handedness on corpus callosum morphology: A meta-analysis. Psychobiology 23(3), 240–247 (1995)
5. Fletcher, P.T., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on Lie groups. In: IEEE CVPR. pp. 95–101 (2003)
6. Jupp, P.E., Kent, J.T.: Fitting smooth paths to spherical data. Applied Statistics 36(1), 34–46 (1987)
7. Kendall, D.G.: Shape manifolds, Procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society 16, 18–121 (1984)
8. Klassen, E., Srivastava, A., Mio, W., Joshi, S.: Analysis of planar shapes using geodesic paths on shape spaces. IEEE PAMI 26(3), 372–383 (2004)
9. Michor, P.W., Mumford, D.: Riemannian geometries on spaces of plane curves. J. Eur. Math. Soc. 8, 1–48 (2006)
10. Miller, M.: Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. NeuroImage 23, S19–S33 (2004)
11. Shi, X., Styner, M., Lieberman, J., Ibrahim, J., Lin, W., Zhu, H.: Intrinsic regression models for manifold-valued data. J Am Stat Assoc 5762, 192–199 (2009)
12. Trouvé, A., Vialard, F.X.: A second-order model for time-dependent data interpolation: Splines on shape spaces. In: MICCAI STIA Workshop (2010)
13. Younes, L.: Computable elastic distances between shapes. SIAM J. Appl. Math. 58, 565–586 (1998)
14. Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. Quarterly of Applied Mathematics (2006)

Varying Coefficient Models for Modeling Diffusion Tensors Along White Matter Bundles

Ying Yuan[1], Hongtu Zhu[1], J. S. Marron[2], John H. Gilmore[3], and Martin Styner[4]

[1] Dept Biostatistics, UNC Chapel Hill
[2] Dept Operations Research and Statistics, UNC Chapel Hill
[3] Dept Psychiatry, UNC Chapel Hill
[4] Dept Psychiatry and Computer Science, UNC Chapel Hill

Abstract. This paper develops a functional data analysis framework to model diffusion tensors along fiber bundles as functional responses with a set of covariates of interest, such as age, diagnostic status and gender. This framework has a wide range of clinical applications including the characterization of normal brain development, the neural bases of neuropsychiatric disorders, and the joint effects of environmental and genetic factors on white matter fiber bundles. A challenging statistical issue is how to appropriately handle diffusion tensors along fiber bundles as functional data in a Riemannian manifold. We propose a statistical model with varying coefficient functions,called VCTF to characterize the dynamic association between functional SPD matrix-valued responses and covariates. We calculate a weighted least squares estimation of the varying coefficient functions under the Log-Euclidean metric in the space of SPD matrices. We also develop a global test statistic to test specific hypotheses about these coefficient functions. Simulated data are further used to examine the finite sample performance of VCTF . We apply our VCTF to study potential gender differences and find statistically significant aspect of the development of diffusion tensors along the right internal capsule tract in a clinical study of neurodevelopment.

1 Introduction

Diffusion Tensor Imaging (DTI), which can track the effective diffusion of water in the human brain *in vivo*, has been widely used to map the microstructure and organization of fiber tracts and to assess the integrity of anatomical connectivity in white matter [1]. In DTI, the degree of diffusivity and the directional dependence of water diffusion in each voxel can be quantified by a 3×3 symmetric positive definite (SPD) matrix, called a diffusion tensor (DT), and its three eigenvalue-eigenvector pairs $\{(\lambda_k, \mathbf{v}_k) : k = 1, 2, 3\}$ with $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Fiber tracts in white matter can be constructed by consecutively connecting the principal directions ($\mathbf{v}_1$) of DTs in adjacent voxels [2]. Therefore, DTs and tensor-derived quantities (e.g., fractional anisotropy (FA)) are distributed along these white matter fiber tracts for each subject. As an illustration, Figure 1 (a)

presents the right internal capsule tract and Figure 1 (b) presents DTs along this tract obtained from 10 subject's, in which each DT is geometrically represented by an ellipsoid. In this representation, the lengths of the semiaxes of the ellipsoid equal the square root of the eigenvalues of a DT, while the eigenvectors define the direction of the three axes. Mathematically, these diffusion tensors along the fiber tract are functionals of SPD matrices. Our research of interest is to statistically model SPD functionals as responses with covariates of interest, such as age and gender, across multiple subjects.

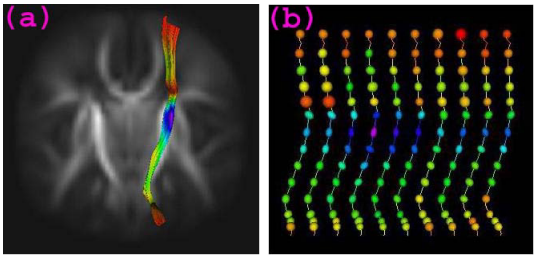


Fig. 1. (a) The right internal capsule tract. (b) The ellipsoidal representation of full tensors on the fiber tract from 10 selected subjects, colored with FA values.

Statistical approaches have been developed for the statistical analysis of tensor-derived quantities along fiber tracts. A tract-based spatial statistics framework was developed to construct local diffusion properties along a white matter skeleton and then perform pointwise hypothesis tests at each grid point of the skeleton [3]. A model-based framework was developed to construct the medial manifolds of fiber tracts and then to test pointwise hypotheses based on diffusion properties along the medial manifolds [4]. However, since these two methods ignore the functional nature of diffusion properties along fiber tracts, they can suffer from low statistical power in detecting interesting features and in exploring variability in tract-based diffusion properties. A functional data analysis framework was used to compare a univariate diffusion property along fiber tracts across two (or more) populations for a single hypothesis test per tract by using functional principal component analysis and the Hotelling T^2 statistic [5]. Their method has two major limitations including only consideration of a univariate diffusion property and the lack of control for other covariates of interest, such as age. To address these two limitations, a functional regression framework was proposed to analyze multiple diffusion properties along fiber tracts as functional responses with a set of covariates of interest, such as age, diagnostic status and gender [6]. An alternative approach, called generalized functional linear models, was developed with a scalar outcome (e.g., diagnostic group) as responses and fiber bundle diffusion properties as varying covariate functions (or functional predictors) [7].

The calculated diffusion properties, which are nonlinear and linear functions of the estimated three eigenvalues of DT containing inherent bias, may be substantially different from the true diffusion properties [8]. Numerical simulations have shown that estimates of the largest eigenvalue in a DT usually overestimate the true value of λ_1 and that estimates of the smallest eigenvalue usually underestimate λ_3. These differences between the estimated and true eigenvalues subsequently bias the estimation of diffusion properties that are calculated from the values of these estimated eigenvalues. The sorting bias is pronounced in three types of degenerate DT including isotropic ($\lambda_1 = \lambda_2 = \lambda_3$), oblate ($\lambda_1 = \lambda_2 > \lambda_3$), or prolate ($\lambda_1 > \lambda_2 = \lambda_3$). Previous studies have shown that a major portion of DTs along fiber tracts are prolate tensors [9], and thus directly comparing these biased diffusion properties along fiber tracts can create 'statistical artifacts' including biased parameter estimates and incorrect test statistics and $p-$values for hypotheses of interest as shown in Section 3.

To avoid these statistical artifacts, it is important to directly analyze estimated DTs along fiber tracts. There are several advantages of comparing the estimated DTs along fiber tracts with covariates. The first one is that the standard weighted least squared estimates of true DTs are almost unbiased [8]. Moreover, as shown in Section 3, directly modeling DTs along fiber tracts as a smooth SPD process allows us to incorporate smoothness constraint to further reduce noise in the estimated DTs along fiber tracts, which subsequently leads to reduced noise in estimated diffusion properties along the fiber tracts. Furthermore, the use of scalar diffusion properties ignores the direction information of DT, and thus it can lose the statistical power in detecting the differences in DT oriented in different directions.

There is a growing interest in the DTI literature in developing statistical methods for direct analysis of DTs in the space of SPD matrices. [10] proposed several parametric models for SPD matrices and derived the distributions of several test statistics for comparing differences between the means of the two (or multiple) groups of SPD matrices. [11] developed a nonparametric estimator for the common density function of a random sample of positive definite matrices. [12] developed a semi-parametric regression model with SPD matrices as responses in a Riemannian manifold and covariates in a Euclidean space. [13] and [14] proposed tensor splines and local constant regressions for interpolating DTI tensor fields based on the Riemannian metric.

In this paper, we propose a varying coefficient model (VCTF) to use varying coefficient functions to characterize the association between fiber bundle diffusion tensors and a set of covariates. This model is different from that in [12] because the former is applicable to DT-valued functional data and consider the within-subject correlations while the latter is only applicable to data with a DT response for each subject. Since the space of SPD matrices is a Riemanian manifold, to the best of our knowledge, our VCTF is the first paper for developing a functional data analysis framework for modeling functional manifold-valued responses with covariates in Euclidean space. To account for the curved nature of the SPD space, we employ the Log-Euclidean metric in [15] and then

use a weighted least squares estimation method based on the geodesic distance under the Log-Euclidean metric to estimate the varying coefficient functions. Furthermore, we develop global test statistics to test hypotheses on the varying coefficient functions and use a resampling method for approximating its p-value.

The rest of the paper is organized as follows. Section 2 presents VCTF and related statistical inference. Section 3 examines the finite sample performance of VCTF via simulation studies. Section 4 illustrates an application of VCTF in a clinical study of neurodevelopment. Section 5 presents concluding remarks.

2 Methodologies

In this section, we present our VCTF for the statistical analysis of DTs along fiber tracts as functional responses with a set of covariates. To compare DTs in populations of DTIs, we use the DTI atlas building followed by atlas fiber tractography and fiber parametrization as described in [5] to extract DTI fibers and establish DTI fiber correspondence across all DTI fiber correspondence across all DTI datasets from different subjects. For the sake of simplicity, we do not include these image processing steps here, which have been discussed in details in [5].

Varying Coefficient Model for Functional SPD data Let $\mathrm{Sym}^+(3)$ and $\mathrm{Sym}(3)$ be, respectively, the set of 3×3 SPD matrices and the set of 3×3 symmetric matrices with real entries. Let $\mathrm{vecs}(C) = (c_{1,1}, c_{2,1}, c_{2,2} \cdots, c_{m_1,1}, \cdots, c_{m_1,m_1})^T$ for any $m_1 \times m_1$ symmetric matrix $C = (c_{k,l})$. Let $\mathrm{Ivecs}(\cdot)$ be the inverse operator of $\mathrm{vecs}(\cdot)$ and (a_l) be a $q \times 1$ vector with the l-th element a_l. Let $C \otimes D$ denote the Kronecker product of two matrices C and D.

Let $x \in [0, L_0]$ be the arc length of any point on a specific fiber bundle relative to a fixed end point of the fiber bundle, where L_0 is the longest arc length on the fiber bundle. For the i-th subject, we measure a diffusion tensor, denoted by $S_i(x_j) \in \mathrm{Sym}^+(3)$, at the arc length $x_j \in [0, L_0]$ for the j-th location grid point on the fiber bundle for $j = 1, \cdots, n_G$ and $i = 1, \cdots, n$, where n_G and n denote the numbers of grid points and subjects, respectively. We consider a varying coefficient model given as follows:

$$\log(S_i(x)) = \mathrm{Ivecs}((\mathbf{z}_i^T \beta_l(x))) + \mathcal{U}_i(x) + \mathcal{E}_i(x) \quad \text{for} \quad i = 1, \cdots, n, \tag{1}$$

where $\log(\cdot)$ denotes the matrix logarithm, $\mathcal{E}_i(x) \in \mathrm{Sym}(3)$ is a 3×3 symmetric matrix of measurement errors, and $\mathcal{U}_i(x) \in \mathrm{Sym}(3)$ characterizes both individual matrix variations from $\mathrm{Ivecs}((\mathbf{z}_i^T \beta_l(x)))$ and the correlation structure between $\log(S_i(x))$ and $\log(S_i(x'))$ for different x and x'. Moreover, $\mathbf{z}_i$ and $\beta_l(x) = (\beta_{1l}(x), \cdots, \beta_{rl}(x))^T$ are, respectively, a $r \times 1$ vector of covariates of interest with $z_{i,1} = 1$ and its associated vector of varying coefficient functions of x for $l = 1, \cdots, 6$. Model (1) can be regarded as a generalization of varying coefficient models, which have been widely studied and developed for longitudinal, time series, and functional data.

Let $\mathrm{SP}(\mu, \Sigma)$ denote a stochastic process with mean $\mu(x)$ and covariance matrix function $\Sigma(x, x')$ for any $x, x' \in [0, L_0]$. It is also assumed that $\mathrm{vecs}(\mathcal{E}_i(x))$ and $\mathrm{vecs}(\mathcal{U}_i(x))$ are independent and respectively, independent and identical copies of $SP(\mathbf{0}, \Sigma_{\mathcal{E}})$ and $SP(\mathbf{0}, \Sigma_{\mathcal{U}})$. Moreover, $\mathrm{vecs}(\mathcal{E}_i(x))$ and $\mathrm{vecs}(\mathcal{E}_i(x'))$ for $x \neq x'$ are assumed to be independent and thus $\Sigma_{\mathcal{E}}(x, x')$ takes the form of $\Sigma_{\mathcal{E}}(x)\mathbf{1}_{x=x'}$.

Weighted Least Squares Estimation To estimate the coefficient functions in $\beta(x) = (\beta_1^T(x), \cdots, \beta_q^T(x))^T$, we develop a weighted least squares estimation method based on an adaptive local polynomial kernel (LPK) smoothing technique [16] and the geodesic distance under the Log-Euclidean metric (see [15] for details). Specifically, using Taylor's expansion, we can expand $\beta_l(x_j)$ at x to obtain $\beta_l(x_j) \approx \beta_l(x) + \dot{\beta}_l(x)(x_j - x)$,

For a fixed bandwidth $h^{(1)}$, we first calculate a weighted least squares estimate of $\beta_l(x)$, denoted by $\hat{\beta}_l(x)$, which minimizes an objective function given by $\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathrm{tr}\{[\log(S_i(x_j)) - \mathrm{Ivecs}((\mathbf{z}_i^T(\beta_l(x) + \dot{\beta}_l(x)(x_j - x))))]^{\otimes 2}\}$, where $K_{h^{(1)}}(\cdot) = K(\cdot/h^{(1)})/h^{(1)}$ is a rescaled kernel function, $K(\cdot)$ be a kernel function, such as the Gaussian and uniform kernels [16] and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector or any matrix $\mathbf{a}$. Then with some calculation, we can have

$$\hat{\beta}_l(x) = (\hat{\beta}_{1k}(x), \cdots, \hat{\beta}_{rl}(x))^T = [I_r \otimes (1, 0)]A_l(x)), \tag{2}$$

where I_r is an $r \times r$ identity matrix, $A_l(x)) = \Sigma(h^{(1)}, x)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)[\mathbf{z}_i \otimes \mathbf{y}_{h^{(1)}}(x_j - x)](\log(S_i(x_j)))_l$ and $\Sigma(h^{(1)}, x) = \sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)[\mathbf{z}_i^{\otimes 2} \otimes \mathbf{y}_{h^{(1)}}(x_j - x)^{\otimes 2}]$ with $\mathbf{y}_{h^{(1)}}(x_j - x) = (1, (x_j - x)/h^{(1)})^T$.

We pool the data from all n subjects and select an estimated bandwidth $h^{(1)}$, denoted by $\hat{h}_e^{(1)}$ by minimizing the cross-validation score given by $(nn_G)^{-1}\sum_{i=1}^{n}$ $\sum_{j=1}^{n_G}\mathrm{tr}\{[\log(S_i(x_j)) - \mathrm{Ivecs}((\mathbf{z}_i^T\hat{\beta}_l(x_j, h^{(1)})^{(-i)}))]^{\otimes 2}\}$, where $\hat{\beta}_l(x, h^{(1)})^{(-i)}$ is the weighted least squares estimator of $\beta_l(x)$ for the bandwidth $h^{(1)}$ based on observed data with the observations from the i-th subject excluded. Finally, by substituting $\hat{h}_e^{(1)}$ into (2), we can obtain an estimate of $\beta_l(x)$, denoted by $\hat{\beta}_{l,e}(x)$. Combining all $\hat{\beta}_{l,e}(x)$ leads to $\hat{\beta}_e(x) = [\hat{\beta}_{1,e}(x), \cdots, \hat{\beta}_{q,e}(x)]$.

Smoothing Individual Functions and Estimating Covariance Matrices To simultaneously construct the individual function $\mathcal{U}_i(x)$, we also employ the local polynomial kernel smoothing technique. Specifically, using Taylor's expansion, we can expand $\mathcal{U}_i(x_j)$ at x to obtain $\mathcal{U}_i(x_j) \approx \mathcal{U}_i(x) + \dot{\mathcal{U}}_i(x)(x_j - x)$. For each fixed x and each bandwidth $h^{(2)}$, we calculate the weighted least square estimator of $\mathcal{U}_i(x)$, denoted as $\hat{\mathcal{U}}_i(x)$, by minimizing an objective function given by $\sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - \mathbf{x})\mathrm{tr}\{[\log(S_i(x_j)) - \mathrm{Ivecs}((\mathbf{z}_i^T\hat{\beta}_{l,e}(x_j))) - \mathcal{U}_i(x) + \dot{\mathcal{U}}_i(x)(x_j - x)]^{\otimes 2}\}$. With some calculation, it can be shown that the weighted least square estimate of $\mathcal{U}_i(x)$, denoted by $\hat{\mathcal{U}}_i(x)$, given by

$$\mathrm{vecs}(\hat{\mathcal{U}}_i(x))^T = \sum_{j=1}^{n_G} \tilde{K}_{h^{(2)}}^0(x_j - x, x)(\mathrm{vecs}(\log(S_i(x_j))) - (\mathbf{z}_i^T\hat{\beta}_{l,e}(x_j)))^T, \tag{3}$$

where $\tilde{K}^0_{h^{(2)}}(x_j - x, x) = \Sigma_1(h^{(2)}, x)^{-1} K_{h^{(2)}}(x_j - x)\mathbf{y}_{h^{(2)}}(x_j - x)$ is the empirical equivalent kernel and $\Sigma_1(h^{(2)}, x) = \sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)\mathbf{y}_{h^{(2)}}(x_j - x)^{\otimes 2}$ with $\mathbf{y}_{h^{(2)}}(x_j - x) = (1, (x_j - x)/h^{(2)})^T$

Let R_i be a matrix with the j-th row $\text{vecs}(\log(S_i(x_j))) - (\mathbf{z}_i^T \hat{\beta}_{l,e}(x_j))^T$ and $\mathcal{S}$ be a $n_G \times n_G$ smoothing matrix with the (i,j)-th element $\tilde{K}^0_{h^{(2)}}(x_j - x_i, x_i)$. We pool the data from all n subjects and select an estimated bandwidth of $h^{(2)}$, denoted as $\hat{h}_e^{(2)}$, by minimizing the generalized cross-validation score given by $n^{-1} \frac{\sum_{i=1}^n \text{tr}\{(R_i - \mathcal{S}R_i)^{\otimes 2}\}}{(1 - n^{-1}\text{tr}(\mathcal{S}))^2}$. Based on $\hat{h}_e^{(2)}$, we can use (3) to estimate $\text{vecs}(\mathcal{U}_i(x))$ and $\mathcal{U}_i(x)$, denoted by $\text{vecs}(\hat{\mathcal{U}}_{i,e}(x))$ and $\hat{\mathcal{U}}_{i,e}(x)$, respectively, for all i. After obtaining $\hat{\mathcal{U}}_{i,e}(x)$, we can estimate the mean function $\mathcal{U}(x)$ and the covariance function $\Sigma_{\mathcal{U}}(x, x')$ by using their empirical counterparts.

We construct a nonparametric estimator of the covariance matrix $\Sigma_{\mathcal{E}}(x, x)$ as follows. Let $\hat{\mathcal{E}}_i(x_j) = \log(S_i(x_j)) - \text{Ivecs}((\mathbf{z}_i^T \hat{\beta}_{l,e}(x_j))) - \hat{\mathcal{U}}_{i,e}(x_j)$ be estimated residuals for $i = 1, \cdots, n$ and $j = 1, \cdots, n_G$. We consider the kernel estimate of $\Sigma_{\mathcal{E}}(x, x)$ given by

$$\hat{\Sigma}_{\mathcal{E}}(x, x) = (n - q)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_G} \frac{K_{h^{(3)}}(x_j - x)\text{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2}}{\sum_{j=1}^{n_G} K_{h^{(3)}}(x_j - x)}. \tag{4}$$

Let $\tilde{\Sigma}_{\mathcal{E}}(x_j, x_j) = (n - q)^{-1} \sum_{i=1}^n \text{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2}$. To select an estimated bandwidth $h^{(3)}$, denoted by $\hat{h}_e^{(3)}$, we minimize the cross-validation score given by $(nn_G)^{-1} \sum_{i=1}^n \sum_{j=1}^{n_G} \text{tr}\{[\text{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2} - \hat{\Sigma}_{\mathcal{E}}(x_j, x_j, h^{(3)})^{(-i)}]^{\otimes 2} \tilde{\Sigma}_{\mathcal{E}}(x_j, x_j)^{-1}]^{\otimes 2}\}$, where $\hat{\Sigma}_{\mathcal{E}}(x, x, h^{(3)})^{(-i)}$ is the weighted least squares estimator of $\hat{\Sigma}_{\mathcal{E}}(x, x)$ based on observed data with the observations from the i-th subject excluded. Based on $\hat{h}_e^{(3)}$, we can use (4) to estimate $\Sigma_{\mathcal{E}}(x, x)$, denoted by $\hat{\Sigma}_{\mathcal{E},e}(x, x)$.

Hypothesis Test In neuroimaging studies, some scientific questions require the comparison of fiber bundle diffusion tensors along fiber bundles across two (or more) diagnostic groups and the assessment of the development of fiber bundle diffusion properties along time. Such questions can often be formulated as linear hypotheses of $\beta(x)$ as follows:

$$H_0 : \mathbf{R}\beta(x) = \mathbf{b}_0(x) \text{ for all } x \text{ vs. } H_1 : \mathbf{R}\beta(x) \neq \mathbf{b}_0(x), \tag{5}$$

where $\mathbf{R}$ is a $t \times 6r$ matrix of full row rank and $\mathbf{b}_0(x)$ is a given $t \times 1$ vector of functions .

We propose both local and global test statistics. The local test statistic can identify the exact location of significant grid point on a specific tract. At a given grid point x_j on a specific tract, we test the local null hypothesis $H_0(x_j) : \mathbf{R}\beta(x_j) = \mathbf{b}_0(x_j)$ against $H_1(x_j) : \mathbf{R}\beta(x_j) \neq \mathbf{b}_0(x_j)$. We use a local test statistic $T_n(x_j)$ defined by

$$T_n(x_j) = n\mathbf{d}(x_j)^T \{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x_j, x_j) \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}^{-1}\mathbf{d}(x_j), \tag{6}$$

where $\hat{\Omega}_Z = n^{-1}\sum_{i=1}^{n}\mathbf{z}_i^{\otimes 2}$ and $\mathbf{d}(x) = \mathbf{R}(\hat{\beta}_o(x)^T - \text{bias}(\hat{\beta}_o(x)^T)) - \mathbf{b}_0(x)$. Following [17], a smaller bandwidth leads to a small value of bias$(\hat{\beta}_o(x))$. Moreover, according to our simulation studies below, we have found that the effect of dropping bias$(\hat{\beta}_o(x))$ is negligible and therefore, we drop it from now on.

We test the null hypothesis $H_0 : \mathbf{R}\beta(x) = \mathbf{b}_0(x)$ for all x using a global test statistic T_n defined by

$$T_n = n\int_0^{L_0} \mathbf{d}(x)^T[\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x,x)\otimes\hat{\Omega}_X^{-1})\mathbf{R}^T]^{-1}\mathbf{d}(x)dx. \tag{7}$$

It follows from Theorem 1 [18] along with the continuous mapping theorem that as both n and n_G converges to infinity, we have T_n converges to some distribution (weighted χ^2). Based on this result, we develop a wild bootstrap method to approximate the p-value of T_n.

3 Simulation Studies

We conducted two sets of Monte Carlo simulations. The first set of simulations was to evaluate the Type I and II error rates of the global test statistic T_n. The second set was to compare the power in detecting the group effect using either whole diffusion tensor or the diffusion properties.

Simulation 1 In the first set of simulations, we evaluated the Type I and II error rates by simulating diffusion tensors along the right internal capsule tract (Figure 1 (a)) according to $S_i(x) = \exp(\text{Ivecs}((\beta_{1l}(x) + \beta_{2l}(x)\times \text{G}_i + \beta_{3l}(x)\times \text{Gage}_i)) + \mathcal{U}_i(x) + \mathcal{E}_i(x))$, where Gage_i and G_i, respectively, denote the gestational age at the scan time and gender of the i-th infant, vecs$(\mathcal{U}_i(x))$ is a Gaussian process with zero mean and covariance matrix $\Sigma_{\mathcal{U}}(x,x')$ and vecs$(\mathcal{E}_i(x))$ is a Gaussian random vector with zero mean and covariance matrix $\Sigma_{\mathcal{E}}(x,x)\mathbf{1}(x = x')$. To mimic imaging data, we used the diffusion tensors along the right internal capsule tract from all the 96 infants in our clinical data to estimate $\hat{\beta}(x)$ of $\beta(x)$ via equation (2), $\hat{\mathcal{U}}(x)$ of $\mathcal{U}(x)$ via equation (3), and $\hat{\mathcal{E}}(x)$ of $\mathcal{E}(x)$ via $\hat{\mathcal{E}}(x) = \log(S_i(x)) - \text{Ivecs}((\hat{\beta}_l(x)^T\mathbf{z})) - \hat{\mathcal{U}}(x)$. We fixed all the parameters at their values obtained from our clinical data, except that we assumed $(\beta_{31}(x),\cdots,\beta_{36}(x)) = c(\hat{\beta}_{31}(x),\cdots,\hat{\beta}_{36}(x))$, where c is a scalar specified below and $(\hat{\beta}_{31}(x),\cdots,\hat{\beta}_{36}(x))$ were estimators obtained from our clinical data. Figure 2 displays the estimated diffusion tensors using VCTF method when $c = 1$. Note that the method did an excellent job of recovering ground truth.

In neuroimaging studies, some scientific questions require the assessment of the development of fiber bundles diffusion tensors across time. In this simulation study, the questions were formulated as the hypotheses test $H_0 : \beta_{31}(x) = \cdots = \beta_{36}(x) = 0$ for all x along the right internal capsule tract against $H_1 : \beta_{3l}(x) \neq 0$ for at least one x on the tract for some $l = 1,\cdots,6$. We first assumed $c = 0$ to assess the Type I error rates for the global test statistic T_n, and then we assumed $c = .2, .4, .6$, and $.8$ to examine the Type II error rates for T_n at different effect

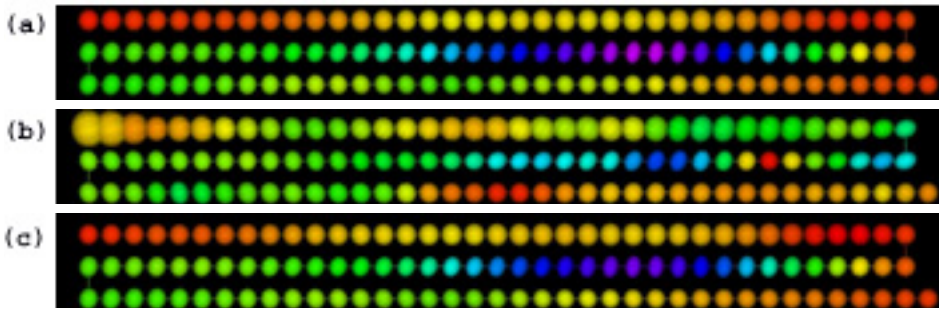


Fig. 2. Ellipsoidal representations of the true (a), simulated (b) and estimated (c) diffusion tensors along the the right internal capsule tract, colored with FA values.

sizes. To evaluate the Type I and II error rates at different sample sizes, we let $n = 96$. The values of gender and gestational age were set the same as all the 96 infants in our clinical study. Note that the number of grid points on the right internal capsule equals $n_G = 112$ for both cases.

We applied the VCTF procedure to the simulated diffusion tensors. Particularly, we approximated the p-value of T_n using the wild bootstrap method described in the hypothesis test section. For each simulation, the significance levels were set at $\alpha = .05$ and .01, and 100 replications were used to estimate the rejection rates. For a fixed α, if the Type I rejection rate is smaller than α, then the test is conservative, whereas if the Type I rejection rate is greater than α, then the test is anticonservative, or liberal.

Figure 3(a) displays the rejection rates for T_n based on the resampling method for sample size 96) and all effect sizes ($c = 0,\ .2, .4, .6,$ or $.8$) at both significance levels ($\alpha = .01$ or $.05$) using full diffusion tensors. The statistical power for rejecting the null hypothesis increases with the effect size and the significance level, which is consistent with our expectation.

Simulation 2 In the second set of simulations, the diffusion tensors along the fiber tract were generated as in the first set of simulation. For each simulated diffusion tensor, we calculated its FA and MD values. Then we applied the VCTF procedure for multiple measures to the simulated values of FA, MD, joint FA and MD, respectively, and then tested the significance of the gestational age effect. It is observed from Fig. 3 (a)-(c) that the statistical power is much higher when we use the whole tensor instead of FA and MD. The power is a little higher at small effect size when we use the joint FA and MD values instead of full diffusion tensors while at higher effect size, they almost have the same power (Fig. 3(a) and (d)). However, the Type I error is greater than the .05 significance level when we use the joint FA and MD values (Fig. 3(d)). It means that the test is liberal. As mentioned in Section 1, this is because there is some inherent bias in these calculated diffusion properties.

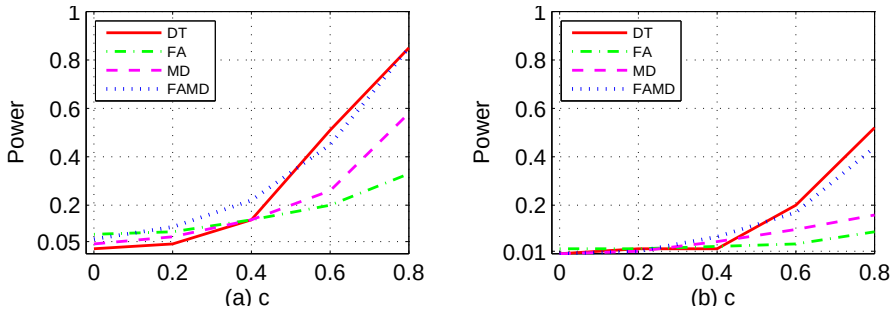


Fig. 3. Simulation study: Type I and Type II error rates. Rejection rates of $\mathbf{T}_n$ based on the resampling method are calculated at five different values of c for sample sizes of 96(solid lines) subjects at the (a) .05 and (b) .01 significance levels using diffusion tensor, FA values, MD values, joint values of FA and MD.

4 A Real Example

We investigate early brain development by using DTI and our VCTF. We consider 96 healthy infants (36 males and 60 females) whose mean gestational age was 245.6 days with SD: 18.5 days (range: 192-270 days). A 3T Allegra head only MR system was used to acquire all the images. The system was equipped with a maximal gradient strength of 40 mT/m and a maximal slew rate of 400 mT/(m·msec). The DTI images were obtained by using a single shot EPI DTI sequence (TR/TE=5400/73 msec) with eddy current compensation. The six non-collinear directions at the b-value of 1000 s/mm^2 with a reference scan (b=0) was applied. To improve the signal-to-noise ratio of the images, a total of five scans were acquired and averaged. A weighted least square estimation method [1] was used to construct the diffusion tensors. Then DTI atlas building followed by atlas based tractography procedure was employed to process all 96 DTI datasets. A nonlinear fluid deformation based highdimensional, unbiased atlas computation method is used to carry out a large deformation non-linear registration [19]. We chose the right internal capsule tract to illustrate the applicability of our method. Diffusion tensors were extracted along this fiber tract for all the 96 infants [5].

In this study, we have two specific aims. The first one is to compare diffusion tensors along the selected fiber bundle across the male and female groups and thus illuminate the gender effect on the development of these fiber bundle diffusion tensors. The second one is to delineate the development of fiber bundle diffusion tensor across the gestational age effect. To statistically test the effects, we applied our VCTF to diffusion tensors along the fiber tract. For the selected tract, we fitted the VCTF model (1) to the diffusion tensors from all 96 subjects, in which $\mathbf{z} = (1, \text{gender}, \text{Gage})^T$. Then, we used equation (2) to estimate the functional coefficients $\beta(x)$. For the hypothesis testing, we constructed the global test statistic T_n via equation (7) to test the gender and age effects for the diffusion tensors. The p value of T_n was approximated using the resampling

method with $G = 1,0000$ replications. Furthermore, we fitted VCTF model to FA, MD,joint FA and MD, each of three eigenvalues and three eigenvalues together values along the fiber tract.

In this study, we statistically test the effects of gender and gestational age on the diffusion tensors along the right internal capsule tract. To test the gender effect, we calculated the local test statistics $T_n(x_j)$ and their corresponding p values across all grid points on the right internal capsule tract. It is observed from Figure 4 (a) that most grid points do not have significant $-\log_{10}(p)$ values, which are less than 1.3. Then, we also computed the p-value of the global test statistic $\mathbf{T}_n$, $p = .335$ indicating no gender effect. It is observed from Figure 4 (b) that the $-\log_{10}(p)$ values of $T_n(x_j)$ for testing the gestational age effect at some grid points of the right tail are greater than 1.5 while a very high significant gestational age effect was found with the $p-$ value of the global test statistic, $p = .01$. This indicates that diffusion tensors along the right internal capsule tract do not differ significantly between male and female groups but are significantly associated with the gestational age.We picked a grid point with the significant p value of $T_n(x)$ and observed that the diffusion tensors become more spherical with the gestational age (Figure 5). This indicates a decreasing pattern of diffusion.

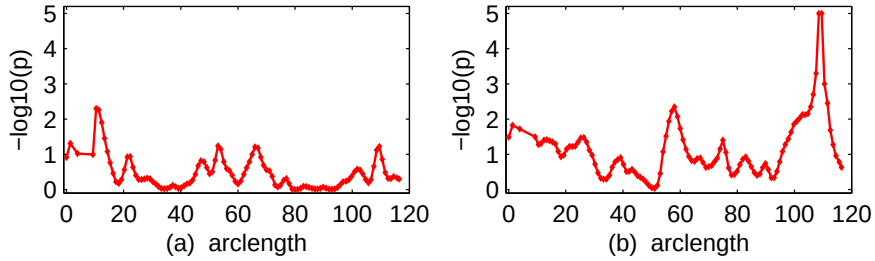


Fig. 4. The $-\log_{10}(p)$ values of test statistics $T_n(x_j)$ for testing gender (a) or gestational age (b) effect of diffusion tensors on the right internal capsule tract.

5 Discussion

We have developed VCTF methods for diffusion tensors along fiber tracts in the Riemannian manifold of SPD matrices under the log-Euclidean metric. From the application end, VCTF is demonstrated in a clinical study of neurodevelopment for revealing the complex inhomogeneous spatiotemporal maturation patterns as the apparent changes in fiber bundle diffusion tensors. We have shown that this novel statistical tool leads to new findings in our clinical applications.

Another commonly used metric on the Riemannian manifold of SPD matrices is the Riemannian metric. In contrast, some operations, e.g. average or interpolation of a set of tensors under the Log-Euclidean and Riemannian metrics, are

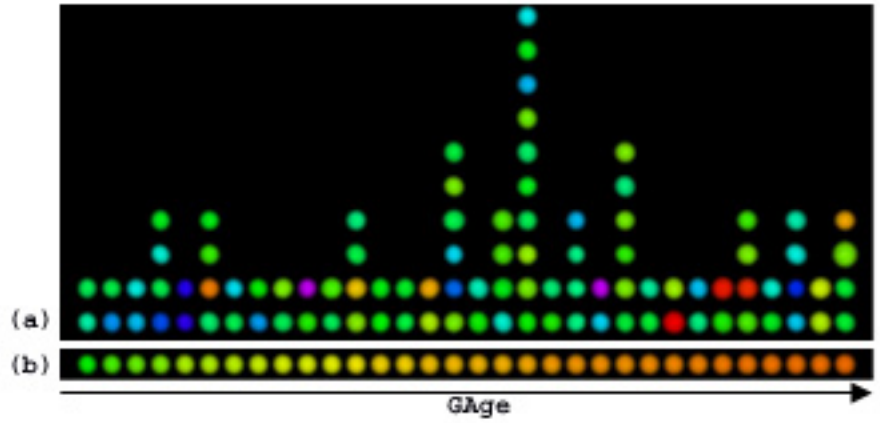


Fig. 5. The ellipsoidal representations of (a) Raw diffusion tensor of all 96 infants and (b) smoothed diffusion tensors changing with the gestational age at one point on the right internal capsule tract with significant gestational age effect.

theoretically and practically very similar [15]. Moreover, some statistical methods based on the two metrics have very similar results. However, the Riemannian metric is affine invariant. Affine invariance is a desirable feature for imaging processing, e.g. segmentation. In this scenario, as shown in [13], the method using the Riemannian metric outperformed the method using the Log-Euclidean metric. So it is interesting to develop VCTF under the Riemannian metric and then compare the statistical powers of detecting group differences under these two different metrics.

References

1. P. J. Basser, J. Mattiello, and D. LeBihan. Estimation of the effective self- diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance Ser. B*, 103:247–254, 1994.
2. P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using dt-mri data. *Magnetic Resonance in Medicine*, 44:625–632, 2000.
3. S. M. Smith, Jenkinson M., H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews, and T. E. Behrens. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31:1487–1505, 2006.
4. P. A. Yushkevich, H. Zhang, T.J. Simon, and J. C. Gee. Structure-specific statistical mapping of white matter tracts. *Neuroimage*, 41:448–461, 2008.
5. C. B. Goodlett, P. T. Fletcher, J. H. Gilmore, and G. Gerig. Group analysis of dti fiber tract statistics with application to neurodevelopment. *NeuroImage*, 45:S133–S142, 2009.
6. H. T. Zhu, M. Styner, N. S. Tang, Z. X. Liu, W. L. Lin, and J.H. Gilmore. Frats: Functional regression analysis of dti tract statistics. *IEEE Transactions on Medical Imaging*, 29:1039–1049, 2010.

7. J. Goldsmith, J. Feder, C. M. Crainiceanu, B. Caffo, and D. Reich. Penalized functional regression. Technical report, Working Paper 2010-204, Johns Hopkins University Dept. of Biostatistics., 2010.
8. H. T. Zhu, H. P. Zhang, J. G. Ibrahim, and B. G. Peterson. Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion). *Journal of the American Statistical Association*, 102:1085–1102, 2007.
9. H. T. Zhu, D. Xu, R. Amir, X. Hao, H. Zhang, K. Alayar, B. Ravi, and B. Peterson. A statistical framework for the classification of tensor morphologies in diffusion tensor images. *Magnetic Resonance Imaging*, 24:569–582, 2006.
10. A. Schwartzman, W. Mascarenhas, and J. E. Taylor. Inference for eigenvalues and eigenvectors of gaussian symmetric matrices. *Ann. Statist.*, 36:1423–1431, 2008.
11. P. T. Kim and D. S. Richards. *Deconvolution density estimation on spaces of positive definite symmetric matrices.* IMS Lecture Notes Monograph Series. A Festschrift of Tom Hettmansperger, 2010.
12. H. T. Zhu, Y. S. Cheng, J. G. Ibrahim, Y. M. Li, C. Hall, and W. L. Lin. Intrinsic regression models for positive definitive matrices with applications in diffusion tensor images. *Journal of the American Statistical Association*, 104:1203–1212, 2009.
13. A. Barmpoutis, B. C. Vemuri, T. M. Shepherd, and J. R. Forder. Tensor splines for interpolation and approximation of dt-mri with applications to segmentation of isolated rat hippocampi. *IEEE Transactions on Medical Imaging*, 26:1537–1546, 2007.
14. B. C. Davis, E. Bullitt, P. T. Fletcher, and S. Joshi. Population shape regression from random design data. *IEEE 11th International Conference on Computer Vision*, pages 1–7, 2007.
15. V. Arsigny. Processing data in lie groups: An algebraic approach. application to non-linear registration and diffusion tensor mri. Ph.D. dissertation, 2006.
16. J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London, 1996.
17. Jianqing Fan and Wenyang Zhang. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.*, 27(4):715–731, 2000.
18. H. T. Zhu, R. Z. Li, and L. L. Kong. Multivariate varying coefficient models for functional responses. Technical report, University of North Carolina at Chapel Hill, 2010.
19. S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:S151–S160, 2004.

Matching and Classification of Images Using The Space of Image Graphs

Dohyung Seo, Jeffrey Ho, and Baba C. Vemuri

Dept of CISE, University of Florida , Gainsville FL, 32611, USA

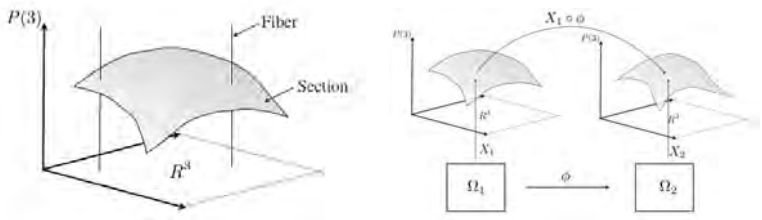
Abstract. This paper proposes a geometric approach for comparing tensor-valued images (tensor fields) that is based on the idea of matching intrinsically low-dimensional shapes embedded in a higher-dimensional ambient space. More specifically, instead of regarding the tensor fields as tensor-valued functions defined on a given (image) domain, we consider their image graphs. These tensorial image graphs can naturally be regarded as submanifolds (shapes) in an ambient space that is the cartesian product of their domain and the space of tensors. With this viewpoint, comparisons between tensor fields can naturally be formulated as comparisons between their corresponding shapes, and an intrinsic comparison measure can be developed based on matching these low-dimensional shapes. The proposed approach offers great conceptual clarity and transparency, and thorny issues such as parametric invariance and symmetric registration can be handled effortlessly in this novel framework. Furthermore, we show that the resulting variational framework can be satisfactorily optimized using a gradient descent-based method, and the computed similarities can be used as the affinity measures in a supervised learning framework to yield competitive results on challenging classification problems. In particular, experimental results have shown that the proposed approach is capable of producing impressive results on several classification problems using the OASIS image database, which include classifying the MR brain images of Alzheimer's disease patients.

1 Introduction

A central application of computational anatomy is to quantify the difference between normal anatomy and pathology. In many application problems, this frequently requires comparisons between 3D volumetric images or shapes extracted from these images. In particular, the difference is often characterized via a similarity measure computed by first registering the two different images/shapes followed by an integration step that sums over the measured difference at corresponding pixels/voxels or surfels (surface elements). Due to its simplicity, this $\mathbf{L}^2$-based approach is very popular in the literature, especially for comparing images and shapes. However, for many applications, tensor fields derived from the scalar-valued images usually contain more useful information that have frequently been under-utilized until now, and for these applications, comparisons between tensor fields are important and fundamental. The examples include the

metric tensor for shape comparison, deformation tensor for morphometry [1], the diffusivity tensor [2] for diffusion MRI and many others. Therefore, there is an increasing need for developing a sound and principled approach for comparing tensor fields, and in this paper we will present one such approach.

The approach taken in this paper is entirely geometrical and it is based on the fundamental idea of comparing the intrinsic shapes of the tensor fields. More precisely, suppose $\mathbf{T}_i$, and $\mathbf{T}_j$ are $\mathbf{P}(n)$-valued tensor fields defined over an image domain Ω considered as a subset in $\mathbb{R}^n$, where $\mathbf{P}(n)$ denotes the space of symmetric positive-definite $n \times n$ matrices (Figure 1). The tensor fields can naturally be regarded as submanifolds in the product space $\Omega \times \mathbf{P}(n)$. Once we have equipped $\Omega \times \mathbf{P}(n)$ with a Riemannian metric, the geometries of $\mathbf{T}_i$, and $\mathbf{T}_j$ are then naturally defined using the induced metrics. In other words, each tensor field is considered as a kind of high-dimensional (tensorial) image graph [5]. For a grey-scale image, its image graph is a submanifold in $\mathbb{R}^3$; however, for a tensor field, its tensorial image graph is contained in a the (nonlinear) ambient space that typically has dimension greater than three. With this particular viewpoint, comparisons between tensor fields can be formulated intrinsically as comparisons between their corresponding tensorial image graphs, and the method of [4] can be generalized directly to this higher-dimensional context by computing the registration between two shapes embedded in the ambient space.



(a) (b)

Fig. 1. Left: The tensorial image graph of a tensor field defined on $\mathbb{R}^3$. The ambient space is the cartesian product $\mathbb{R}^3 \times \mathbf{P}(3)$. **Right:** Comparing shapes in $\mathbb{R}^3 \times \mathbf{P}(3)$ can be realized via a registration map X defined between the two shapes.

The approach outlined above offers several important advantages over the relatively straightforward $\mathbf{L}^2$-based approach. First, it is conceptually transparent and by comparing the shapes (instead of tensor fields) directly on the tensorial image graphs, the issue of parametrization invariance becomes irrelevant and the complicated procedures of finding a parametrically invariant representation [2] can be completely bypassed. Second, the issue of symmetric comparison and registration [3] is naturally incorporated through the use of intrinsic volume forms defined on the shapes. Third, while the mathematics is more involved when compared with the $\mathbf{L}^2$-based methods, it is still tractable and we have developed a gradient descent-based method to efficiently optimize the objective function. Fi-

nally, the proposed approach is also flexible in that it permits different metrics to define different shapes for a given tensor field. For example, the ambient space $\Omega \times \mathbf{P}(n)$ can be equipped with several different metrics (for example, the usual Frobenius metric and affine-invariant metric on $\mathbf{P}(n)$), and our approach can easily accommodate these different choices.

For validation, we apply the proposed method to classify brain MRI from different population groups contained in the OASIS database [10] which includes the challenging problem of classifying MR brain scans of patients suffering from the alzheimer's disease and that of normal healthy old adults. The MR brain images are first converted into their corresponding tensor representations, and pairwise similarities between the resulting tensor fields are computed using the proposed method. Nonlinear dimensionality reduction of the data is achieved using a diffusion map [11] with the pairwise similarities as the basic affinity measures, and the final classification step is carried out efficiently and accurately in a low-dimensional Euclidean space. In terms of better and stronger classification results, the experimental results reported in section 4 validate the two novel points advocated in this paper that the tensor field comparisons are fundamental and important in many applications, and a sound and principled approach to tensor field comparison can form the basis of algorithmic solutions to challenging classification problems.

2 Cost Function for Shapes Matching

Shapes can be represented as R^n-dimensional point clouds, triangular meshes, or parametric or implicit surfaces. In this paper, shapes are considered as parametric surfaces. For instance, surfaces in 3-D are represented parameterically by : $f : \Omega \to R^3$,where Ω is a domain. If we represent images as image graphs in order to represent images as shapes, the desired parameterization is $f : (u, v) \to (x, y, I(x, y))$ for 2-D gray scale images or $f : (u, v) \to (x, y, P(3)(x, y))$ for 2-D 3-by-3 symmetric positive definite (SPD) tensor-valued images. More precisely, an image can be considered as a section of a fiber bundle [5]. Then for a pair of shapes (S_1, S_2), a matching problem can be formulated as follows :

$$D((S_1, S_2), \gamma) = \min_{\gamma} \int Dist(S1, S2 \circ \gamma)\sqrt{\kappa} d\Omega, \tag{1}$$

where S_1 and S_2 share common domain Ω. On this stage we assume that shapes are already globally registered. In Eq.(2), $Dist(S_1, S_2)$ is a point-wise distance defined in the ambient space between two shapes and $\sqrt{\kappa} d\Omega$ is $volume form$ defined as $||df_p(u) \times df_p(v)||$ where $f : (u, v) \to p$ and $\sqrt{\kappa}$ is invariant under re-parametrization.

For shape comparison, $D((S_1, S_2), \gamma)$ is required to be invariant under re-parametrization, and symmetric between S_1 and S_2. Recently, [2] introduces the notion of $q - map$ as a novel representation for shapes in $\mathbb{R}^3$ with the aim of achieving re-parametrization invariance of $D((S_1, S_2), \gamma)$. However if the ambient space is not Euclidean, e.g., $\mathbf{P}(3)$, it is not clear how to complete the centering

step in the construction of the $q-map$ in [2]. A similar cost function as Eq.(2) was proposed earlier in [4] which is designed for matching 3-D face meshes. However, the proposed matching algorithm is not symmetric and nor is it designed for matching tensorial image graphs.

In this paper, we introduce a novel symmetric matching framework for low-dimensional shapes embedded in an ambient space X that generally has dimension greater than three. Such shapes will be represented by their parameterizations with domains in $\mathbb{R}^2$ or $\mathbb{R}^3$. Let S_1 and S_2 be two such surfaces with parameterizations f_1 and f_2 defined on two domains Ω_1 and Ω_2 respectively. We will use the following cost function to define a similarity measure between these two shapes

$$E((S_1,S_2),\gamma)=\min_{\gamma}\int (Dist(f_1,f_2\circ\gamma))^2(\sqrt{\kappa_1}+\sqrt{\kappa_2(\Omega_2\circ\gamma)}J_\gamma)d\Omega, \qquad (2)$$

where $\sqrt{\kappa_1}d\Omega_1$ and $\sqrt{\kappa_2}d\Omega_2$ are the pull-backs of the volume form on S_1, S_2 under f_1 and f_2, respectively. $\Omega=\Omega_1$, and J_γ is the determinant of the Jacobian of γ. The details of derivation of Eq.(2) is given in the appendix.

2.1 Tensorial Image Graphs

In this section, we apply our matching framework to tensor-valued images, specifically 3-by-3 3-D SPD tensor-valued images. For this purpose, each tensor-valued image is represented as a section of a fiber bundle [5] with the map $\mathbf{X}:\mathbb{R}^3\rightarrow \mathbb{R}^3\times\mathbf{P}(3)$ or $\mathbf{X}:(u,v,w)\rightarrow(x(u,v,w),y(u,v,w),z(u,v,w),I(x,y,z))$. In this map, $x=u$,$y=v$,$z=w$ and $I(x,y,z)\in\mathbf{P}(3)$ at each voxel. This parametrization was introduced by Gur et. al [5] to achieve anisotropic smoothing of 2-D DTI. In this appplication, the volume form is $\sqrt{\kappa}dudvdw$ where κ is the determinant of

$$\mathbf{K}=\begin{pmatrix} <\mathbf{X_u},\mathbf{X_u}> & <\mathbf{X_u},\mathbf{X_v}> & <\mathbf{X_u},\mathbf{X_w}> \\ <\mathbf{X_u},\mathbf{X_v}> & <\mathbf{X_v},\mathbf{X_v}> & <\mathbf{X_v},\mathbf{X_w}> \\ <\mathbf{X_u},\mathbf{X_w}> & <\mathbf{X_v},\mathbf{X_w}> & <\mathbf{X_w},\mathbf{X_w}> \end{pmatrix}, \qquad (3)$$

or

$$\mathbf{K}=\begin{pmatrix} \lambda+Tr((I^{-1}I_x)^2) & Tr((I^{-1}I_x)(I^{-1}I_y)) & Tr((I^{-1}I_x)(I^{-1}I_z)) \\ Tr((I^{-1}I_x)(I^{-1}I_y)) & \lambda+Tr((I^{-1}I_y)^2) & Tr((I^{-1}I_y)(I^{-1}I_z)) \\ Tr((I^{-1}I_x)(I^{-1}I_z)) & Tr((I^{-1}I_y)(I^{-1}I_z)) & \lambda+Tr((I^{-1}I_y)^2) \end{pmatrix}, \qquad (4)$$

and the metric in the ambient space is given by:

$$\mathbf{G}=\begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mathbf{P} \end{pmatrix}, \qquad (5)$$

where $\mathbf{P}$ is the metric in $\mathbf{P}(3)$ space. We will call the shapes (submanifolds) represented by the parametrization $\mathbf{X}$ as the tensorial image graph of the tensor-valued images I. Representing the tensorial image graphs S_1 and S_2 by (x_1,y_1,z_1,I_1)

and (x_2, y_2, z_2, I_2) respectively, we define $Dist((S_1, S_2), \gamma)$ in Eq.(2) by:

$$Dist((S_1, S_2), \gamma) = \lambda((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2) + dist(I_1, I_2), \quad (6)$$

and

$$dist(I_1, I_2) = Trace(\log(I_1^{-1/2} R^{\dagger}(I_2 \circ \gamma) R I_1^{-1/2})^2). \quad (7)$$

Eq.(7) is voxel-wise Riemannian distance between tensors [6], with R the matrix used to re-orient the tensors (as they undergo a registration). When we choose Ω as the common domain, $(x_1, y_1, z_1) = (u, v, w)$ and $(x_2, y_2, z_2) = (u + U(u, v, w), v + V(u, v, w), w + W(u, v, w))$, where U, V, W are the components of the deformation field. In Eq.(7), R is reoriented with respect to the deformation vector field γ (or U, V, W) at each voxel. Reorientation of tensors must be carried out during optimization to and the reorientation transformation is derived from the deformation field [7, 8]. Also, recall that $Trace((\log(A)^2) = \sum_{i=1}^{3} (\log(\Lambda_i))^2$, where Λ_i's are the eigenvalues of A.

To smooth the vector field, (U, V, W), we add the regularization term to Eq.(2) :

$$E_{reg} = \int |\nabla U|^2 + |\nabla V|^2 + |\nabla W|^2 d\Omega, \quad (8)$$

and the final cost function, $E_{tot}((S_1, S_2), \gamma)$ is given as:

$$E_{tot}((S_1, S_2), \gamma) = (1 - \alpha) E((S_1, S_2), \gamma) + \alpha E_{reg}, \quad (9)$$

where α is a small positive scalar.

To efficiently solve the resulting optimization problem, we first discretize the cost function to get,

$$E_{tot}(\mathbf{U}, \mathbf{V}, \mathbf{W}) = \sum_{i \in \Omega} [(1 - \alpha) Dist(U_i, V_i, W_j)(\sqrt{\kappa_{1i}} + \sqrt{\kappa_{2i}} J_{\gamma i}) + \alpha \sum_{k=\{x,y,z\}} (U_{ki}^2 + V_{ki}^2 + W_{ki}^2)] \quad (10)$$

where $U_i, V_i, W_j, U_{ki}, V_{ki}, W_{ki}$ are the values of U, V, W at the given discrete points. We optimize the above cost function with respect to U, V, W using non-linear conjugate gradient (NCG) method following [9]. To simplify the steps, we evaluate the gradient vector and the Hessian matrix of Eq.(10) with fixed $(\sqrt{\kappa_{1i}} + \sqrt{\kappa_{2i}} J_{\gamma i})$ in each NCG iteration, and in turn, $(\sqrt{\kappa_{1i}} + \sqrt{\kappa_{2i}} J_{\gamma i})$ is computed using the most recently updated deformation vector field, (U_i, V_i, W_i).

3 Application : Classification of MRI data

To apply our matching framework to classification of tensor-valued image data set, we first convert the input images into their associated tensor-valued images (tensor fields) and compute pairwise matching. In the second step, we use the L2 norms of the deformation vector fields, $d(S_i, S_j)$ between the two tensorial

image graphs S_i and S_j to build a data graph with Gaussian weights $e^{-d(S_i,S_j)/\epsilon}$ according to [11]. The corresponding Markov matrix is used to compute the diffusion maps [11], which provides a dimensionality reduction of the input image data. The nearest neighbor classifier is then used for classification using the diffusion distances in the low-dimensional feature space.

3.1 Data Preparation

To create 3-D tensor images from a 3D gray-scale images, we used the OASIS MRI database [10]. The images in the database are the MR human brain scans of subjects aged between 18-96 of 416. Each image has a resolution of $208\times176\times176$ voxels. Our goal is to classify the MR image data into different age groups, and we have chosen the ventricle as the ROI (region of interest) as it captures the part of brain showing the most significant difference across ages. Fig.2 shows ROIs from sagittal and longitudinal MR slices acquired from brains of 18, 43 and 81 year old subjects respectively.

Our tensor-valued images are the first fundamental forms (metric tensors) of image graphs of the 3-D intensity values with map $\mathbf{f} : R^3 \to R^3 \times R$,

$$\begin{pmatrix} <\mathbf{f_u},\mathbf{f_u}> & <\mathbf{f_u},\mathbf{f_v}> & <\mathbf{f_u},\mathbf{f_w}> \\ <\mathbf{f_u},\mathbf{f_v}> & <\mathbf{f_v},\mathbf{f_v}> & <\mathbf{f_v},\mathbf{f_w}> \\ <\mathbf{f_u},\mathbf{f_w}> & <\mathbf{f_v},\mathbf{f_w}> & <\mathbf{f_w},\mathbf{f_w}> \end{pmatrix}, \tag{11}$$

and we consider these tensor fields as tensorial image graphs. Fig.3 shows the tensor-valued images from subjects in Fig.2 according to Eq.11.

3.2 Diffusion Map and Diffusion Distance for Classification

In this paper, we represent the SPD tensor-valued images as sections of a fiber bundle, therefore we need to find a meaningful geometric description of the space of sections for classification purposes and diffusion maps can generate efficient representation of desired geometric structures based on the diffusion processes.[11] Once we build a graph with Gaussian weights $e^{-d(S_i,S_j)/\epsilon}$ and construct the corresponding Markov matrix, $\mathbf{M}$, then the family of diffusion maps $\{\Psi_t\}_{t\in N}$ and diffusion distance $D_t(S_i, S_j)$ are defined as follows :

$$\mathbf{\Psi_t}(\mathbf{S_i}) = \begin{pmatrix} \Lambda_1^t \Psi_1(S_i) \\ \Lambda_2^t \Psi_2(S_i) \\ \vdots \\ \Lambda_s^t \Psi_s(S_i) \end{pmatrix}, \tag{12}$$

and

$$D_t(S_i, S_j) = \left(\sum_{l=1}^{s} \Lambda_l^{2t} (\Psi_l(S_i) - \Psi_l(S_j))^2 \right)^{1/2}, \tag{13}$$

where $\{\Lambda_l\}_{l\geq 0}$ and $\{\Psi_l\}_{l\geq 0}$ are eigenvalues and eigenvectors of $\mathbf{M}$ respectively such that $1 = \Lambda_0 > |\Lambda_1| \geq |\Lambda_2| \geq \ldots$ and $\mathbf{M}\Psi_l = \Lambda_l\Psi_l$ (Ref.), and

$$s = max\{l \in N \text{ such that} |\Lambda_l|^t > \delta|\Lambda_1|^t\}. \tag{14}$$

The diffusion map Ψ_t embeds all data in the set, $\{S_i\}$ into the Euclidean space $\mathbf{R}^s$ and D_t reflects the connectivity in the graph of the data in the set, $\{S_i\}$ defining the Euclidean distance in $\mathbf{R}^s$: points in the set $\{S_i\}$ are closer if they are highly connected in the graph.

We set $d(S_i, S_j)$ as L2 norm of deformation vector fields, or $\int(|U|^2 + |V|^2 + |W|^2)d\Omega$, after matching and use the diffusion distance as feature of nearest neighbor classifier in the low dimensional space.

Fig.4 shows Ψ_1 vs. Ψ_2 plots for classifications between groups. The details of classification results are reported in next section.

4 Experimental Results

In this section, we report the experimental results on classifications of MR brain images from the OASIS data set. We divided the subjects into three groups : the young group with age below 40, the old group with age 60 or above and the middle-aged group between 41-60. In these experiments, we use a four-fold cross-validation and leave-one-out validation to determine the classification scores. In the four-fold cross-validation, the subgroups are randomly selected 50 times and the maximum, minimum and average classification scores together with the variances are reported in the first three columns in Table 1. We also test our method to classify Alzheimer's disease (AD). We take 70 subjects from old age group and in the subgroup, 35 of them are diagnosed as AD and rest of them are control, and we use four-fold and leave-one-out validation tests within the subgroup[13]. The classifier used in the reduced dimension is the nearest-neighbor classifier, and the diffusion distance is used as distance measure. The criterion for determining the dimension of the diffusion map in these experiments is given by $\delta = 0.07$ in Eq.(14) with $t = 1$. The metric used for the ambient space $\mathbb{R}^3 \times \mathbf{P}(3)$ is the product metric of the Euclidean metric on $\mathbb{R}^3$ and the affine-invariant metric on $\mathbf{P}(3)$. Furthermore, tensor reorientation [7],[8] is applied to reorient the tensors after transformation. We note that for three different sets of classification, the classification rates are uniformly high.

In the second set of experiments, we test the effects of using tensor field (compared with only scalar field) and the choice of different metric on $\mathbf{P}(3)$. First, we change the metric on $\mathbf{P}(3)$ from the affine-invariant metric to the Frobenius norm (i.e., $\mathbf{L}^2$-norm), and four-fold cross-validation results between young and old age group are reported in the first two columns in Table 2. In these two columns, we test the effect of tensor reorientation on classifications, and the results show that for Frobenius norm, the effect is generally small. In the third column, the images are represented by their image graphs and the shape comparisons are carried out in the ambient space $\mathbb{R}^3 \times \mathbb{R}$ instead of $\mathbb{R}^3 \times \mathbf{P}(3)$. In this experiment, distances between image graphs are the Riemannian distances

and $\lambda = 0.000001$ in Eq.(6) and $\alpha = 0.02$ in Eq.(9). We remark that the result clearly demonstrates that the classification result using only scalar-valued images (image graphs in $\mathbb{R}^3$) is inferior to the one using tensor fields and the affine-invariant metric on $\mathbf{P}(3)$ provide superior classification result compared with the Frobenius metric.

In Table 3, we show the comparisons between our method and several previously published classification results on the OASIS database. [12] uses the deformation tensor field (computed from registering the image to an atlas) as the main feature for each image. Submanifold of each age group is constructed from the training samples and the geodesic distances between subjects and the submanifolds are used as the main discriminative feature for classification. In [13], alternatively, histograms of deformation vector fields have been used as features, and the CAVIAR method proposed in [13] takes a adaboost-like approach to integrate the results from a collection of weak classifiers into a strong classification result. We remark that our method compared favorably with these methods in terms of classification rates, and in particular, for the more challenging problem of classifying brain images of Alzheimer's disease patients, our method demonstrates a small but real improvement over these two methods.

Table 1. Scores of leave-one-out and four-fold cross-validation test of four subgroups randomly selected 50 times. The metric used for the ambient space $\mathbb{R}^3 \times \mathbf{P}(3)$ is the product metric of the Euclidean metric on $\mathbb{R}^3$ and the affine-invariant metric on $\mathbf{P}(3)$.

	Old vs. Young	Old vs. Middle	Middle vs. Young	AD vs. Control
Maximum	100%	100%	100%	100%
Minimum	97.72%	90.77%	87.27%	75.0%
Average	99.25 %	97.6%	94.36 %	94.87%
Standard deviation	0.8384%	2.46%	3.09%	5.55%
Leave-one-out	99.15 %	98.46 %	96.36%	95.32%

Table 2. Scores of leave-one-out and four-fold cross-validation test of four subgroups randomly selected 50 times between Young and Old samples using Frobenius metric on $\mathbf{P}(3)$ with and without tensor re-orientation. Last column gives the classification result of using only gray-scale images.

Methods	Frobenius with reo.	Frobenius without reo.	Image Graph
Maximum	100%	100%	98.86%
Minimum	88%	88.63%	88.63%
Average	94.96 %	96.52%	94.29 %
Standard deviation	2.57%	2.15%	2.29%
Leave-one-out validation	97.03%	98.01%	94.89%

Table 3. Comparison of classification scores with 4-fold validation between classification methods

	Old vs. Young	Old vs. Middle	Middle vs. Young	AD vs. Control
Image Graphs	99.25%	97.6%	94.36%	94.87%
CAVIAR [13]	99.14 %	98.36 %	97.76%	88.0%
Adaboost [13]	98.75 %	96.80 %	96.0%	90.25 %
Submanifold projection [12]	96.43%	90.23%	84.32%	88.57%
Nearest Neighbor in PCA [12]	92.43%	87.74 %	78.42 %	84.29 %

5 Conclusion

We have proposed a novel geometric approach for comparing tensor-valued images (tensor fields) that is based on the simple idea of matching the low-dimensional tensorial image graphs formed by the tensor fields. Our framework provides a registration method that is both symmetric and invariant under different parametrization, and the resulting cost function can be satisfactorily optimized using a gradient descent-based method. We have reported four different classification experiments using the OASIS image database, and our method has produced results that are in par or exceeding the current state-of-the-art results. In particular, our experiments have shown that tensor fields do indeed contain subtle information that can be useful for challenging classification problems, and the experimental results have demonstrated that the proposed method, although more elaborat and involved compared with $\mathbf{L}^2$-based method, is able to access and utilize this information to obtain good classification results.

References

1. N. Lepore, C. Brun, Y. Chou, M. Chiang, R. A. Dutton, K. M. Hayashi, E. Luders, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, P. M. Thompson, "Generalized Tensor-Based Morphometry of HIV/AIDS Using Multicariate Statistics on Deformation Tensors", IEEE Transations on Medical Imaging, 2008.
2. S. Kurtek, E. Klassen, Z. Ding and A. Srivastava, "A novel riemannian framework for shape analysis of 3D objects", *Proc.* CVPR pp. 1625-1632, 2010.
3. H. D. Tagare, D. Groisser and O. Skrinjar, "Symmetric Non-rigid Registration : A Geometric Theory an Some Numerical Techniques" , J Math Imaging Vis pp. 61-88, 2009.
4. N. Litke, M. Droske, M. Rumpf and P. Schröder, "An Image Processing Approach to Surface Matching", in "Symposium on Geometry Processing", pp. 207-216, 2005.
5. Y. Gur and N. Sochen, "Coordinate-based diffusion over the space of symmetric positive-definite matrices", in "Visualization and Processing of Tensor FIelds - Advances and Perspectives", pp 325-340, Springer-Verlag, Berlin Heidelberg, 2009 .
6. M. Moakher and P. G. Batchelor, "Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization", in "Visualization and Processing of Tensor FIelds", pp285-298 , Springer-Verlag, New York, 2006.

7. G. Cheng, B. C. Vemuri, M. Parekh, P. R. Carney, and T. H. Mareci, "Non-rigid Registration of HARDI Data Represented by a Field of Gaussian Mixtures", In LNCS (Springer) Proceedings of MICCAI09: Int. Conf. on Medical Image Computing and Computer Assisted Intervention, 2009.
8. D. C. Alexander, C. Pierpaoli, P. J. Basser, and J. C. Gee, "Spatial Transformations of diffusion tensor magnetic resonance images", IEEE *Trans. Med. Imag. 20* 11 pp. 1131-1139, 2001.
9. S.H. Lai and B. C. Vemuri, "Reliable and Efficient Computation of Optical Flow", IJCV 29(2), pp. 87-105, 1998.
10. D. S. Marcus, T. H. Wang, J. parker, J. G. Csernansky, J. C. Moris, R. L. Buckner "Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults". Journal of Cognitive Neuroscience (2007)
11. R. R. Coifman and S. Lafon, "Diffusion Maps", Appl. Comput. Harmon. Anal. 21 pp. 5-30, 2006
12. Y. Xie, B. C. Vemuri, and J. Ho, "Statistical Analysis of Tensor Fields ", In Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2010.
13. T. Chen, A. Rangarajan, and B. C. Vemuri, "CAVIAR: Classification via Aggregated Regression and Its Application in Classifying the OASIS Brain Database", In IEEE International Symposium on Biomedical Imaging, pp. 1337-1340, 2010.

Appendix. Cost Function for Symmetric Matching

If we have two parameterized shapes, S_1 and S_2 with domain Ω_1 and Ω_2, then the cost function for symmetric matching problem is formulated as following :

$$E((S_1, S_2), \phi, \psi) = \min_{\phi} \int_{\Omega_1} Dist(S_1, S_2 \circ \phi)\sqrt{\kappa_1} d\Omega_1 + \min_{\psi} \int_{\Omega_2} Dist(S_2, S_1 \circ \psi)\sqrt{\kappa_2} d\Omega_2, \tag{A-1}$$

where $\phi : \Omega_1 \to \Omega_2$ and $\psi : \Omega_2 \to \Omega_1$, and $\sqrt{\kappa} d\Omega$ is $volume form$. If The second term in the left hand side of Eq(A-1) can be rewritten as following :

$$\int_{\Omega_1} Dist(S_2 \circ \psi^{-1}, S_1)\sqrt{\kappa_2(\Omega_2 \circ \psi^{-1})} J_{\psi^{-1}} d\Omega_1, \tag{A-2}$$

and if we require that Eq(A-1) is symmetric matching, ψ^{-1} should be ϕ and $J_{\psi^{-1}} = J_\phi$ which is determinant of Jacobian such as

$$Det \begin{pmatrix} \frac{\partial u_2}{\partial u_1} & \frac{\partial u_2}{\partial v_1} & \frac{\partial u_2}{\partial w_1} \\ \frac{\partial v_2}{\partial u_1} & \frac{\partial v_2}{\partial v_1} & \frac{\partial v_2}{\partial w_1} \\ \frac{\partial w_2}{\partial u_1} & \frac{\partial w_2}{\partial v_1} & \frac{\partial w_2}{\partial w_1} \end{pmatrix} \tag{A-3}$$

And κ_2 is determinant of $\mathbf{K}_2$ given as following :

$$\mathbf{K}_2 = \begin{pmatrix} < S_{2u_2}, S_{2u_2} > & < S_{2u_2}, S_{2v_2} > & < S_{2u_2}, S_{2w_2} > \\ < S_{2u_2}, S_{2v_2} > & < S_{2v_2}, S_{2v_2} > & < S_{2v_2}, S_{2w_2} > \\ < S_{2u_2}, S_{2w_2} > & < S_{2v_2}, S_{2w_2} > & < S_{2w_2}, S_{2w_2} > \end{pmatrix} \tag{A-4}$$

$$= (J_\phi^{-1})^2 \mathbf{K_2'}, \tag{A-5}$$

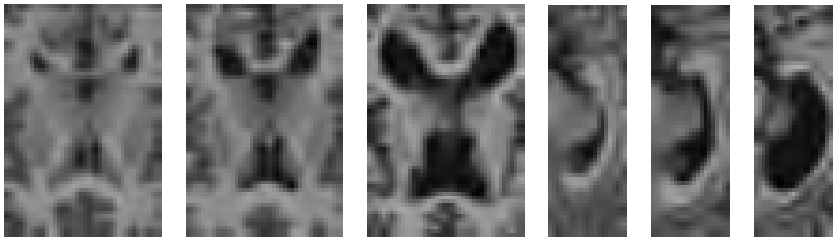
and

$$\mathbf{K_2'} = \begin{pmatrix} < S_{2u_1}, S_{2u_1} > < S_{2u_1}, S_{2v_1} > < S_{2u_1}, S_{2w_1} > \\ < S_{2u_1}, S_{2v_1} > < S_{2v_1}, S_{2v_1} > < S_{2v_1}, S_{2w_1} > \\ < S_{2u_1}, S_{2w_1} > < S_{2v_1}, S_{2w_1} > < S_{2w_1}, S_{2w_1} > \end{pmatrix} \tag{A-6}$$

Finally the cost function is given as following :

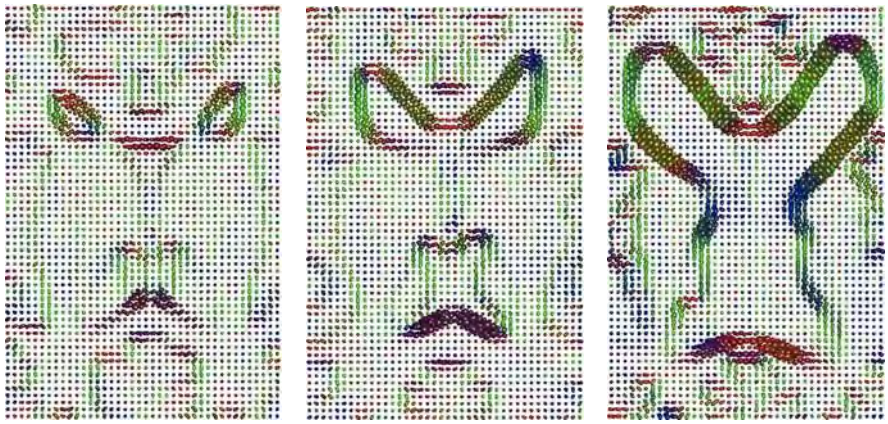
$$E((S_1, S_2), \phi, \psi) = \min_{\phi} \int_{\Omega_1} Dist(S_1, S_2 \circ \phi)(\sqrt{\kappa_1} + \sqrt{\kappa_2'}) d\Omega_1, \tag{A-7}$$

where κ_2' is the determinant of $\mathbf{K_2'}$.



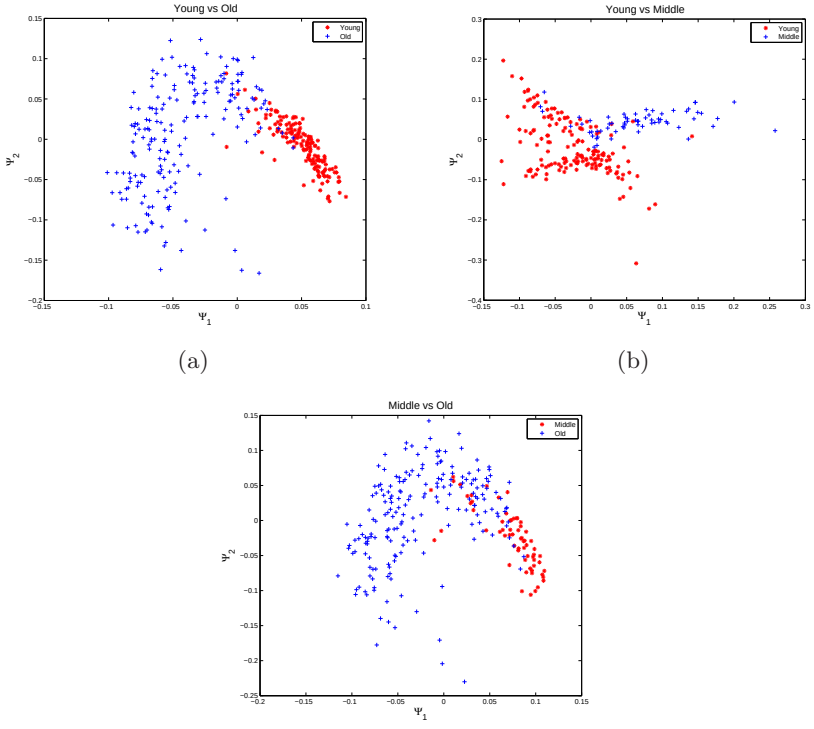
(a) (b) (c) (d) (e) (f)

Fig. 2. Slices of 3-D MRI images. (a)-(c) : Cross-sectional images of ventricles of 18, 43, and 81 years old respectively. (d)-(f): Longitudinal images of ventricles in the same order.



(a) (b) (c)

Fig. 3. Slices of 3-D tensor images of ventricles created by Eq.(11). (a)-(c) : 18, 43, and 81 years old respectively.



(a) (b)

(c)

Fig. 4. 2-D plots of diffusion maps: (a) young vs. old, (b) young vs. middle, and (c) middle vs. old. In each plot, x-axis and y-axis are Ψ_1 and Ψ_2, respectively

Detecting Long Distance Conditional Correlations Between Anatomical Regions Using Gaussian Graphical Models

Stéphanie Allassonnière, Pierre Jolivet and Christophe Giraud *

CMAP Ecole Polytechnique, Route de Saclay, 91128 Palaiseau cedex, France
http://www.cmap.polytechnique.fr

Abstract. The conditional correlation patterns of an anatomical shape may provide some important information on the structure of this shape. We propose to investigate these patterns by Gaussian Graphical Modelling. We design a model which takes into account both local and long-distance dependencies. We provide an algorithm which estimates sparse long-distance conditional correlations, highlighting the most significant ones. The selection procedure is based on a criterion which quantifies the quality of the conditional correlation graph in terms of prediction. The preliminary results on AD versus control population show noticeable differences.

Keywords: Gaussian Graphical Model, long-distance conditional correlation, sparse dependencies, deformation of Alzheimer's disease

1 Introduction

It is quite natural to think that any region of the brain depends on all the other regions at least via some detours. This is equivalent to say that these regions are all correlated to each other. Several studies [10, 12, 5] have tried to design correlation patterns by computing the correlation matrix given by correlations between any two regions (using a PCA decomposition for example). Then, they highlight the most significant or stable ones by thresholding. However, correlations describe the global statistical dependencies between variables, corresponding to both direct and indirect interactions. The *direct* relations between two of these regions are less numerous and harder to capture but they carry some interesting information as well. Focusing on these direct dependencies, called *conditional correlations*, enables to see which areas directly affect the behaviour of a given region and avoids the burden given by the indirect interactions.

We want to study the geometric conditional dependencies of a response given by a real valued signal carried by a discrete grid. A natural approach to estimate these conditional correlations is given by Gaussian Graphical Modelling (GGM)

* The authors thank Johns Hopkins University for the data and Dominique for the cartoon. Support for the JHU group was provided in part by P41-RR015241 (JHU) and U24 RR021382 (MGH) to BIRN, National Center for Research Resources (NIH), P01-AG0568 and P01-AG03991 (WU), and the NSF via DMS-0456253.

[11, 15, 7]. The response is modelled as a Gaussian random vector.The conditional correlations of the response variables are depicted by a graph. Each node represents a variable and an edge in the graph is set between two variables if they are conditionally dependent given all the remaining ones. This corresponds to a non-zero entry of the inverse covariance matrix called the precision matix.

In the context of medical data, we face two problems with Gaussian Graphical Modelling. The first one is the high dimension of the data compared to the low sample size. The second problem is that in this framework, the neighbour points are very likely to be conditionally correlated since the medical phenomenon is usually continuous and these local dependencies can be predominant in the estimated graph possibly hiding any other links. We propose here an algorithm and a selection criterion which address both questions.

The high-dimension-low-sample-size paradigm is a common issue in medical imaging.This usually leads to non accurate and non stable statistical analysis. In our graphical framework, this is equivalent to a high dimensional graph and small number of measures of the random signal. To face this situation, sparse representation is well known to be a very powerful tool in Computer Vision and Pattern recognition [26]. Although we know that the underlying real graph is not sparse, it is interesting to perform a *sparse estimation of its structure* [13, 14]. Such an estimation provides both a stable estimate of the conditional dependencies and a selection of the edges with the highest conditional correlations. The procedure proposed by [15, 13] is based on the LASSO algorithm [21]. In addition, the elastic net algorithm [27] slightly relaxes this sparse constrain providing more accurate and robust-to-noise sparsity patterns [9]. This sparsity constraint is crucial as it prunes the graph keeping a small number of edges.

As noticed above, local dependencies of medical images are predominant while estimating the conditional correlations. We want to go beyond that and focus on long-distance dependencies. To show only long distance dependencies, the authors in [10] post-process their graph merging some nodes and edges when the nodes are closer than a fixed threshold. However, we want to introduce this knowledge as a prior to the estimation. We have brought in the algorithm a neighbourhood prior: knowing that the neighbours are linked, we only look for the other conditional dependencies which are modelled as sparse. This enables to highlight the long distance conditional correlations and to gain in term of statistical accuracy since the estimation of long-distance relations is less affected by the noise and the neighbours. The prior knowledge is modelled as a neighbouring graph provided by the user. Because the neighbouring graph is given by the user, it can take into account geographical closeness or some anatomical proximity given for example by fibres.

The estimation procedure depends on some free parameters. In order to find suitable values, one can either use a cross-validation [13] or a theoretical analysis based on a restrictive condition on the data [14]. Unfortunately these choices are not satisfactory in practice and the final choice is usually done by hand looking at the estimations. To avoid this subjective choice, Giraud [7] proposes, for the usual sparse GGM framework, a criterion which optimises the choice of the graph

depending on its capacity to predict one random variable given the others. We generalise this criterion to the case of neighbourhood prior.

In Computational Anatomy, learning what characterises the population in terms of dependencies as in [19] for Alzheimer's disease (AD) is an important aspect of the analysis of the pathology. Different but complementary methods have been used to analyse populations for instance Deformation Based-Morphometry (DBM) [3], Tensor Based-Morphometry [25] or Feature Based-Morphometry [22]. The general idea is to compute some statistics from a sample of template-to-subject deformations. Statistical analysis can also be done directly on the deformations [3, 6, 2] or by modelling the generation of images [18, 1] showing the admissible deformations in a population. Although the global deformations give an interesting interpretation, we may want to know what parts of the shape are deformed jointly. Our proposed algorithm based on LASSO or elastic net has been tested on a training set of high dimension data representing some deformations of a template hippocampus towards 101 patients' hippocampi. The results show some correlated regions when dealing with deformations which suggest that this structure has regions which move together. Moreover, the population is divided into two groups, controls and patients suffering from AD. The comparison of the estimated graphs of the two groups show some noticeable differences which may be a new characterisation of AD.

2 Gaussian Graphical Models

Let us consider p points on a given shape that will compose the nodes of the graph. On these points, we observe n random responses such as a discretisation (with p points) of a quantification of template-to-subject deformations. The p nodes of the graph are thus identified to p random variables denoted $(X_1, ..., X_p)$, this vector is assumed to be distributed as a multivariate Gaussian $\mathcal{N}_p(0, \Sigma)$ (data are re-centred if necessary). The graph G_Σ of conditional dependencies is defined as follows: there exists an edge between nodes a and b if and only if the variables X_a and X_b are dependent given all the remaining variables. This will be denoted $a \overset{G_\Sigma}{\sim} b$. To illustrate the notion of conditional dependency, let us give a toy example illustrated in Fig. 1. The traffic jam intensity and the number of snowmen in town are correlated due to snowstorm. But conditionally on the height of snow, the number of snowmen is independent of the intensity of traffic jams. This corresponds to the graph in Fig 1 with only two edges. In particular, there is no edge between the traffic jam and snowmen variables.

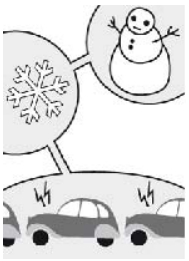


Fig. 1. Conditionally on the height of snow, the number of snowmen is independent of the intensity of traffic jams. This is represented by a two edges graph.

2.1 Estimation processes

The estimation is done by a regression as presented in [15]. Let X_a be the a^{th} node of the graph. The goal is to estimate the matrix θ such that: $X_a =$

$\sum_{b \neq a} \theta_{a,b} X_b + \varepsilon_a$ where ε_a is assumed to follow a normal distribution with zero-mean and variance $1/(\Sigma^{-1})_{a,a}$. An important property of this modelling is that $\theta_{a,b} = -\frac{K_{a,b}}{K_{a,a}}$ where $K = \Sigma^{-1}$ is the precision matrix. Therefore $\theta_{a,b} = 0$ is equivalent to $a \overset{G_\Sigma}{\not\sim} b$. Let $\theta_{.,b}$ be the p-dimensional vector of $(\theta_{a,b})_{1 \leq a \leq p}$ and X_{-a} the $(p-1)$-dimensional random vector $(X_1, ..., X_{a-1}, X_{a+1}, ..., X_p)$. The estimation of θ is done by minimising an energy (which is actually the negative penalised log likelihood of the model) given by two terms. The first one is the l^2 distance between the observations and their estimation (coming from the Gaussian distribution of the noise). The second term is a prior on the parameters θ to estimate. In the case we consider, the prior is a Laplacian prior with parameter λ. This corresponds to an l^1 penalty on the energy side. The energy to minimise is therefore the following:

$$\forall\, b \in \{1, ..., p\}, \quad \hat{\theta}_{.,b} = \underset{\theta_{.,b}}{\operatorname{argmin}} \|X_b - X_{-b}\theta_{.,b}\|_2^2 + \lambda \|\theta_{.,b}\|_1 \,. \tag{1}$$

This regression technique is known as the *LASSO* algorithm [21]. Introducing a penalty on the estimated parameter θ enables to estimate it. Indeed, θ may be of large dimension which prevents from using a small sample size n. Reducing the estimation to more likely or expected matrices, enables to handle this typical case in medical imaging.

The minimum argument of this energy is sparse thanks to the l^1 penalty. Besides, it provides a stable estimation in terms of prediction of a variable given the other ones. Including this constraint on the candidate matrices enables to really focus on the most important conditional correlations. Therefore predicting one variable from the others is not much dependent on the training set.

However, one may want to relax slightly this constraint by balancing the effect of this l^1 penalty by an l^2 one which has the opposite behaviour. Instead of looking for sparse matrix with big coefficients, it tends to spread to all elements in the matrix. This is known as the elastic net algorithm [27] which provides more accurate sparsity patterns [9]. The energy is:

$$\forall\, b \in \{1, ..., p\}, \quad \hat{\theta}_{.,b} = \underset{\theta_{.,b}}{\operatorname{argmin}} \|X_b - X_{-b}\theta_{.,b}\|_2^2 + \lambda \|\theta_{.,b}\|_1 + \gamma \|\theta_{.,b}\|_2^2 \,. \tag{2}$$

These two algorithms take each node independently one after the other. This has a drawback considering that we would like to have a symmetric relation between nodes. Minimising either energy above does not guarantee that $\hat{\theta}_{a,b} \neq 0$ at the same time as $\hat{\theta}_{b,a} \neq 0$. The construction of the graph is therefore not straightforward. Two natural choices appear to enforce the symmetry [15]. On the one hand, we can say that only one of $\hat{\theta}_{a,b}$ or $\hat{\theta}_{b,a}$ needs to be non-zero to set an edge in the graph between a and b. This yields $a \overset{\widehat{G}_\Sigma}{\sim} b$ if and only if $\hat{\theta}_{a,b} \neq 0$ or $\hat{\theta}_{b,a} \neq 0$. On the other hand, we can prefer a harder decision process by saying that only one of the $\hat{\theta}_{a,b}$ or $\hat{\theta}_{b,a}$ needs to be zero to remove the edge between a and b in the graph. This leads to $a \overset{\widehat{G}_\Sigma}{\sim} b$ if and only if $\hat{\theta}_{a,b} \neq 0$ and $\hat{\theta}_{b,a} \neq 0$. Both will be tested in the sequel.

2.2 Penalised criterion to choose among a family of estimated graphs

In the two previous energies (1) and (2), there are two parameters which have to be chosen. This may be very difficult and data depending. A classical choice is to use cross-validation. Giraud in [7] suggests instead to set a criterion so that after computing a family of possible graphs with different parameters, it enables to select the best one with respect to this criterion which aims at answering the question of how good the network is to predict one variable from the others.

To define the criterion crit(.) we need to introduce a few notations. We associate to any graph G in $\widehat{\mathcal{G}}$, the $p \times p$ matrix

$$\widehat{\theta}^G = \arg\min_{\theta \in \Theta_G} \sum_{1 \le i \le n, 1 \le a \le p} [X - X\theta]^2_{i,a} \,,$$

where Θ_G is the set of $p \times p$ matrices θ such that $\theta_{a,b}$ is non-zero if and only if there is an edge between a and b in G. Then, we define the criterion crit(G) by

$$\text{crit}(G) = \sum_{a=1}^{p} \left[\|X_a - \sum_b X_b \widehat{\theta}^G_{a,b}\|^2_n \left(1 + \frac{\text{pen}[d_a(G)]}{n - d_a(G)}\right)\right], \tag{3}$$

where the penalty function is defined by

$$\text{pen}(d) = 1.1 \times \frac{n-d}{n-d-1} \,\text{EDKhi}\left[d+1, n-d-1, \left(\binom{p-1}{d}(d+1)^2\right)^{-1}\right] . \tag{4}$$

The function EDKhi$[d, N, .]$ is the inverse of the function

$$\text{DKhi}[d, N, .] \;:\; x \mapsto \mathbb{P}\left(F_{d+2,N} \ge \frac{x}{d+2}\right) - \frac{x}{d}\,\mathbb{P}\left(F_{d,N+2} \ge \frac{N+2}{Nd}x\right),$$

where $F_{d,N}$ denotes a Fisher random variable with d and N degrees of freedom. See [7] and [4] for more details and more explanations on this penalty. This Criterion (3) is implemented in a `R` library GGMselect [8].

3 Non local Gaussian Graphical Models

3.1 Introduction of G_0

As noticed above, the neighbouring points - neighbour nodes of the graph- are very likely to be conditionally correlated. We want to put more attention onto the other correlations -which will be called long-distance ones in the sequel since they do not affect the neighbours. To this purpose, we have introduced in the estimation a neighbouring graph G_0 which carries the neighbour nodes of all the graph nodes. We assume that there exist correlations between these points but we are not estimating them rather looking for the other ones. There are two reasons for that. On the one hand, we are not interested in the local correlations since they appear to be obvious. The long-distance ones however may reveal

some important behaviour which may be characteristic from the population we are studying. These non obvious relations between regions may show some non random effect on the shape. This will be illustrated in the experiments.

On the other hand, the value of the penalty is directly related to the complexity of the families of possible candidate graphs. Shrinking the families to those of graphs containing G_0 reduces their complexity and therefore the penalty. The effect it that more edges appear and therefore the long-distance ones.

The way we introduce G_0 in the estimation is as follows. The estimation of the conditional correlations is done in the orthogonal space of the neighbouring graph G_0. This leads to replace the random variable X_a by

$$X_a - X_{m_a}(X_{m_a}^T X_{m_a})^{-1} X_{m_a}^T X_a \,, \tag{5}$$

where X^T is the transposition and X_{m_a} is the matrix defined as follow: if we denote X the $n \times p$ matrix of all the data and m_a the list of neighbours of node a in G_0, then $X_{m_a} = X(., m_a)$ is of dimension $n \times card(m_a)$.

This orthogonality constraint may have to be relaxed since the projection may lead to ill-conditioned matrices. Moreover, it will also enable to capture some edges with small projection onto the orthogonal of G_0 which may appear stronger since we know that the local dependencies are more likely to be the strongest and summarise the main information. This can be express in terms of introducing a small ridge (driven by a new parameter γ_0 chosen via the criterion). Denoting Id the identity matrix, this yields

$$X_a - X_{m_a}(X_{m_a}^T X_{m_a} + \gamma_0 Id)^{-1} X_{m_a}^T X_a \,. \tag{6}$$

3.2 Selection of graph: Change in the penalty

When we introduce an a priori graph G_0, we shall change the selection criterion to take into account this a priori knowledge. To fit with this situation, the Criterion (3) remains unchanged, but the penalty (4) is replaced by

$$\mathrm{pen}_{G_0}(a,d) = 1.1 \times \frac{n-d}{n-d-1}\,\mathrm{EDKhi}\left[d+1, n-d-1, \left(\binom{p-d_0[a]-1}{d-d_0[a]}(d-d_0[a]+1)^2\right)^{-1}\right] \tag{7}$$

where $d_0[a]$ is the degree of the node a in the graph G_0. This correction reflects the change of complexity induced by the graph G_0. It ensures a control of the prediction error similar to Theorem 1 in [7]. For further details, we refer to the Appendix. We have therefore modified GGMselect in two ways (which is included in the new update of the package). If a prior is given, we change the data so that the new random variable is given by eq. (5) or (6). Then we changed the penalty criterion for eq. (7).

4 Experiments

We have tested this model on meshed surfaces of hippocampi from Johns Hopkins University [16]. The data base contains $n = 101$ vectors of dimension $p = 713$

corresponding to the vertices of the graph. A template surface of a hippocampus was mapped onto each subject surface using Large Diffeomorphic Deformation Metric Mapping [17]. Prior rigid body registration has been done on each subject. In this common coordinate system, the log of the absolute value of the Jacobian determinant of this deformation was discretised onto the template meshed surface. This leads to 101 vectors with real value signal which correspond to scalar deformation fields over the surface. The subjects belong to three sub-populations. The first 57 were control patients. The 32 next were suffering from mild Alzheimer's disease or semantic dementia (denoted mild AD) and the 12 last were in a late stage of the disease (denote late AD in the sequel). Some examples of the training set are presented in Fig. 2.

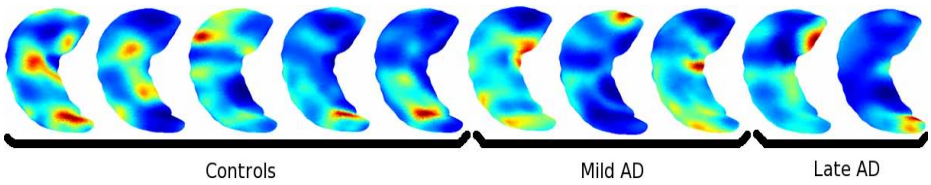


Fig. 2. Examples of the training set. The colour depends on the intensity of the Jacobian of the deformation. Blue means a contraction and red dilatation. The intensity itself is not important but rather its relative value with respect to the others.

4.1 Gaussian Graphical estimation without neighbour prior

The results of the graph estimations without any neighborhood prior are presented in the four graphs on the left hand side of Fig. 3. The four plots correspond respectively to the solution of eq.(1) with both choice of symmetrisation ("or" and "and" resp.) and of eq.(2) again with both symmetrisation processes. For each of these minimisation, we present the result with the lowest criterion. All these graph present some similarities. Indeed, they all find the conditional correlations between close nodes of the graph with respect to the euclidean distance. This is what we expect since the deformations are smooth. Therefore, the deformation at one point is close to the deformation on its neighbours leading to a correlation between all nodes but only local conditional correlations. However, one of the graphs (left one, with a slightly lower criterion) suggests that they should be some conditional correlation not only between neighbour nodes. As we can see, not all the part of the surface are linked through long edges. However, this does not appear in the three other graphs which makes them uncertain.

4.2 Non local GGM estimation

To overcome the problem of the predominance of the conditional correlations between close nodes, we have introduced a neighbouring graph G_0. Two examples of G_0s are presented in Fig. 3 (right). The estimated graphs using our proposed

algorithm are presented in Fig. 4 where the top row corresponds to the first G_0 of Fig. 3 and the bottom row to the second one. All the eight estimations are presented: lasso (with the two possible symmetrisations) using the orthogonal projection and the ridge regularisation and next, in the same order, the four estimation with the elastic net energy. The orthogonal projection may appear too strong since it requires to invert the neighbouring matrix (cf eq. (5)). This yields no graph estimated with the lasso minimisation process with a strong symmetrisation constraint on the first line.

We can notice that they almost all carry some long-distance correlations. In particular, four of them show some common patterns pointing the same connected areas (row 1, columns 2,4,5,6). The estimation remains stable when adding or removing some edges in G_0. This is shown in the bottom row of Fig 4. The results are very similar. Noticing that the orthogonal of this new G_0 is smaller (the maximum degree of this G_0 is 7 whereas the first one was 3) the resulting graphs look alike with a little less edges. In this case, the estimation of the first graph is possible but presents very few connections. As well as for the first G_0, the lasso with a relaxed projection constrains provides very interesting graphs and the elastic net does not requires this relaxation. Since some graphs appear very dense, we removed the edges between close points w.r.t. the Euclidean distance. We did this pruning for the marked graph (see below for meaning of this mark) of Fig. 4. The result is presented in Fig. 5. Thanks to this representation, the long-distance conditional correlation appear clearer and can be easier interpreted. Indeed, this suggest that there is a strong conditional correlation between the bottom part of the head of the hippocampus and two regions of the body. This means that the hippocampus is subjected to deformations that are not random. When the bottom area of the head is deformed so are the two local parts of the body. Note that the conditional correlation can be positive or negative which correspond to same and opposite behaviour respectively.

Thanks to the criterion we can point the graph which, among the ones we estimate, has the best power for prediction purposes. This graph is marked with a red star. This graph looks indeed very interesting showing long-distance dependencies that are not trivial. The power of this criterion we propose is that

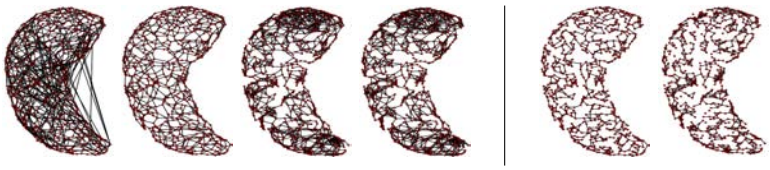


Fig. 3. Left: Estimated graphs using the lasso algorithm and both symmetrisations (called "or" and "and" in the sequel) and the elastic net algorithm with both symmetrisations as well ("or" and "and" resp.). The results show the predominance of the neighbours conditional correlations except for the first estimation where long-distance conditional correlation appear. Right: two examples of possible G_0s.

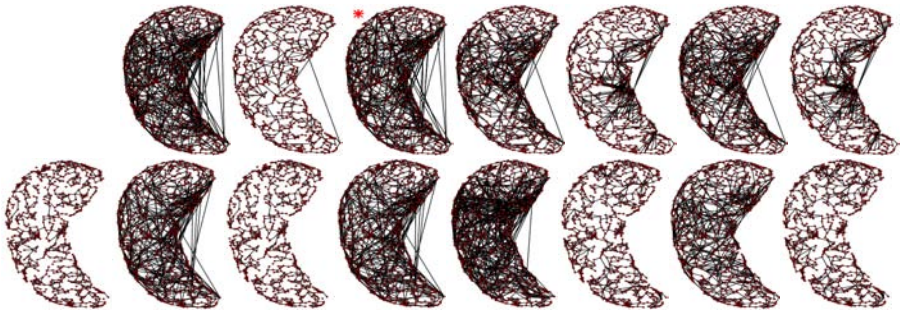


Fig. 4. Estimated graphs with the eight methods using the two neighbouring graphs presented in Fig. 3 on the top and bottom line resp. From left to right, each method with both symmetrisations ("or" and "and" resp.): lasso with the orthogonal projection, lasso with a relaxation on the projection, elastic net with projection, elastic net with relaxed projection. The projection constraint may be too strong preventing from computing the estimation (top left). In almost all graphs, long-distance dependencies appear and are stable with respect to the prior neighbours.

it can quantify the predictive power of any other graph estimated by any other method. We not only provide some estimation technics which take into account a neighbourhood prior but also a way to compare them to other estimated graphs.

4.3 Clustering of the graph

In order to see the different areas which are conditionally correlated, we use the spectral clustering method which highlights these dependencies. Spectral clustering is a technique based on eigen-properties of a similarity matrix Q that partitions data into disjoint clusters. We compute the p eigenvectors of Q, $(V_i)_{1\leq i\leq p}$. Let the vectors $Y_j = (V_i^j)_{2\leq i\leq p}$ be the contatenation of the j^{th} coordinates of each eigenvector. The clustering is done using a k-mean algorithm on these vectors. The choice of the similarity matrix is given as follows. If two nodes i and j are connected through the graph but not in the neighbouring graph G_0 then $Q(i,j) = 1/Z$. If this condition is not satisfied then $Q(i,j) = f(\|x_i - x_j\|)/Z$ where Z is a renormalisation constant, x_i is the pose of node i and f is the Gaussian density function with fixed variance (chosen with respect to the spreading of the data, here 50 whereas $\max(x_i - x_j)^2 = 7185$). $Q(i,i)$ is set so that the sum of the i^{th} row is 0. The result using 5 clusters is presented in Fig. 5 where each node has a colour corresponding to its cluster. We can clearly see that the regions that are in the same cluster (same colour) correspond to areas that share a link in the pruned graph (Left of Fig. 5). This shows that several connected parts of the shape (red, black, green and dark blue) seem to have conditionally independent behaviour whereas some nodes (cyan) are clearly related. This result confirms the segmentation in [24].

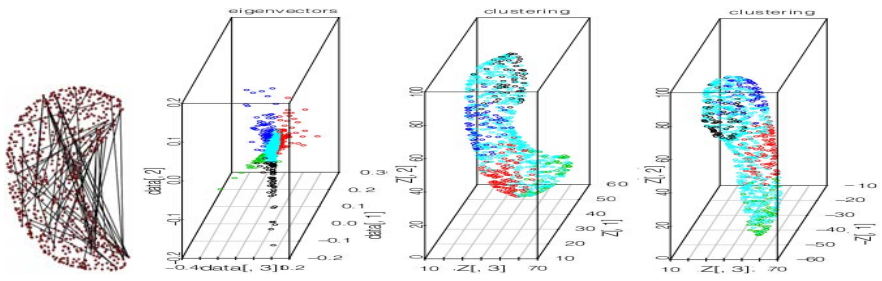


Fig. 5. Best graph analysed. Left: the edges between close nodes have been removed for better visualisation. Right: k-mean clustering.

4.4 Population different

Since we are provided with three sub-groups, it is interesting to see whether they carry differences with respect to conditional correlations. However, since the late AD group is pretty small, we only cluster it into two groups: controls versus AD (both mild and late together). The goal is to see if the disease changes the kind of conditional correlated deformations that are present in a population. This is actually what Fig 6 tends to show. The deformations of the hippocampus in a control population are not completely random since some important long distance conditional correlations appear. This reduces the safe deformations to a subset where these joint behaviour occurs. However, the AD population seems to have less long-distance conditional correlation. This would suggests that the disease affects the hippocampus by removing the direct dependencies of different regions. This effect cannot be seen when no neighbourhood obvious correlations are taken into account (graphs on the left of Fig. 6). However, when introducing the projection onto the orthogonal of a neighbouring graph (the first one presented above), the differences appear and are stable with respect to the estimation process (the criterions are of the same range for these four tests).

5 Conclusion

This paper presents a new way of analysing populations of shapes in terms of random graphs which carry sparse conditional correlations between areas of a shape. We have introduced a neighbourhood prior. It stabilises the estimation and highlights the long-distance conditional correlations which are the non obvious ones. This neighbouring graph is given by the user allowing for non trivial closeness notion as some anatomical ones. The results on the deformed hippocampi reveal some important conditional correlations between particular sub-regions which are stable along the estimation processes as well as with respect to a change in the neighbouring graph. Moreover, it emphasises differences between the control subjects which have more long-distance edges than the AD group. Provided with a larger database would help confirming this first trend. In

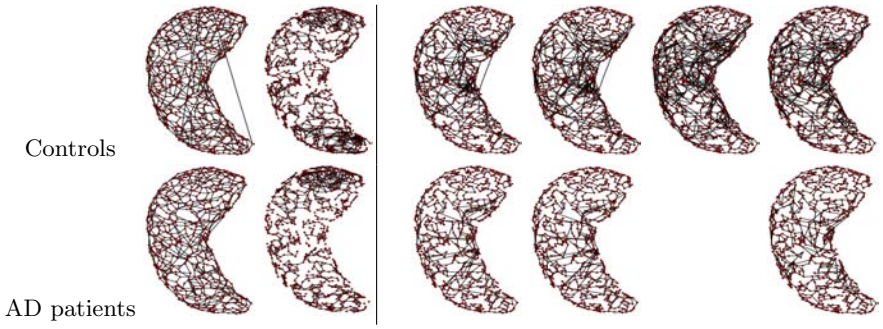


Fig. 6. Comparison of the estimated graphs from the two sub-groups. Left: estimations without a prior on the neighbours (lasso "and" and elastic net "or" resp.). Right: estimations using the first G_0 presented above (lasso "and" with orthogonal projection and the ridge relaxation, elastic net "or" with projection (when computable) and ridge relaxation resp.). The graph without the prior are very similar and does not enable any discrimination. Introducing the prior knowledge enables to catch the differences: the AD group present less long-distance conditional correlations.

addition, we provide the user with a criterion which quantifies the quality of any graph (not only the one estimated with our algorithm). The estimation has been done here for anatomical shapes but is more general and it would be interested to test this on functional data to catch *direct long-distance* dependencies in the brain connectivity network and compare with for eg [23] and [20].

6 Appendix

The penalty pen_{G_0} is a complexity penalty built from the theory of Giraud [7]. For a given integer D and a graph G_0, we introduce the collection $\mathcal{M}_D(G_0)$ of graphs of degree less than D and containing G_0. A careful inspection of the proof of Theorem 1 in [7] shows that for each node a the penalty $\text{pen}_{G_0}(a,d)$ must ensures the control

$$\sum_{m_a \in \mathcal{M}_0(a)} (|m_a|+1)\text{DKhi}\left(|m_a|+1, n-|m_a|-1, \frac{(n-|m_a|-1)\text{pen}(a,m_a)}{K(n-|m_a|)}\right) \leq \log(n),$$

where $\mathcal{M}_0(a) = \left\{\{b : b \overset{G}{\sim} a\},\ G \in \mathcal{M}_D(G_0)\right\}$.This control is achieved by eq (7).

References

1. Allassonnière, S., Kuhn, E., Trouvé, A.: Bayesian deformable models building via stochastic approximation algorithm: A convergence study. Bernoulli J. 16(3), 641–678 (2010)
2. Allassonnière, S., Younes, L.: A stochastic algorithm for probabilistic independent component analysis. In revision for Annals of Applied Statistics
3. Ashburner, J., et al.: Identifying global anatomical differences: deformation-based morphometry. Human Brain Mapping 6(5-6), 348–357 (1998)

4. Baraud, Y., Giraud, C., Huet, S.: Gaussian model selection with an unknown variance. Annals of Statistics 37(2), 630–672 (2009)
5. Bickel, P., Levina, E.: Covariance regularization by thresholding. Annals of Statistics 36(6), 2577–2604 (2008)
6. Durrleman, S., et al.: Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. Medical Image Analysis 12(5), 626–637 (2008)
7. Giraud, C.: Estimation of gaussian graphs by model selection. Electron. J. Stat. 2, 542–563 (2008)
8. Giraud, C., Huet, S., Verzelen, N.: Graph selection with ggmselect. arXiv:0907.0619
9. Jia, J., Yu, B.: On model selection consistency of the elastic net when p $\gg$ n (2008)
10. Kim, S.G., et al.: Stuctural connectivity via the tensor-based morphometry. Biannual Meeting of Korean Society of Human Brain Mapping (2010)
11. Lauritzen, S.: Graphical models 17 (1996)
12. Lee, H., et al.: Discriminative persistent homology of brain networks (2011)
13. Lee, H., et al.: Sparse brain network using penalized linear regression. SPIE Medical Imaging (2011)
14. Lee, H.e.a.: Sparse brain network recovery under compressed sensing. IEEE Transactions on Medical Imaging (2011)
15. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. Ann. Statist. 34(3), 1436–1462 (2006)
16. Miller, M., et., al.: Morphometry BIRN. collaborative computational anatomy: An MRI morphometry study of the human brain via diffeomorphic metric mapping. Human Brain Mapping 30(7), 2132–2141 (2009)
17. Miller, M.I., Trouvé, A., Younes, L.: On the metrics and Euler-Lagrange equations of computational anatomy. Annual Review of biomedical Engineering 4 (2002)
18. Sabuncu, M., Balci, S.K., Golland, P.: Discovering modes of an image population through mixture modeling. MICCAI LNCS(5242), 381–389 (2008)
19. Seeley, W., et al.: Neurodegenerative diseases target large-scale human brain networks 62(1), 42–52 (2009)
20. Smith, S., et al.: Network modelling methods for fMRI. NeuroImage 54(2), 875–891 (2011)
21. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc B 58 (1), 267–288 (1996)
22. Toewsa, M., et al.: Feature-based morphometry: Discovering group-related anatomical patterns. Neuroimage 49(3), 2318–2327 (2010)
23. Varoquaux, G., et al.: Brain covariance selection: better individual functional connectivity models using population prior. NIPS (2010)
24. Wang, L., et al.: Abnormalities of hippocampal surface structure in very mild dementia of the alzheimer type. Neuroimage 30:52-60 (2006)
25. Wang, Y., Chan, T.F., Toga, A.W., Thompson, P.M.: Shape analysis with multivariate tensor-based morphometry and holomorphic differentials (2009)
26. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T., Yan, S.: Sparse representation for computer vision and pattern recognition. Proceedings of the IEEE (2010)
27. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B (2005)

An Airway Tree-shape Model for Geodesic Airway Branch Labeling

Aasa Feragen[1], Pechin Lo[1], Vladlena Gorbunova[1], Mads Nielsen[1], Asger Dirksen[2], Joseph M. Reinhardt[3], François Lauze[1], Marleen de Bruijne[1] *

[1]Department of Computer Science, University of Copenhagen, Denmark
{aasa, marleen}@diku.dk
[2]Department of Orthopedics and Internal Medicine, University of Copenhagen, Denmark
[3]Department of Biomedical Engineering, University of Iowa, USA
http://image.diku.dk/aasa

Abstract. We present a mathematical airway tree-shape framework where airway trees are compared using geodesic distances. The framework consists of a rigorously defined shape space for treelike shapes, endowed with a metric such that the shape space is a geodesic metric space. This means that the distance between two tree-shapes can be realized as the length of the geodesic, or shortest deformation, connecting the two shapes. By computing geodesics between airway trees, as well as the corresponding airway deformation, we generate airway branch correspondences. Correspondences between an unlabeled airway tree and a set of labeled airway trees are combined with a voting scheme to perform automatic branch labeling of segmented airways from the challenging EXACT'09 test set. In spite of the varying quality of the data, we obtain robust labeling results.

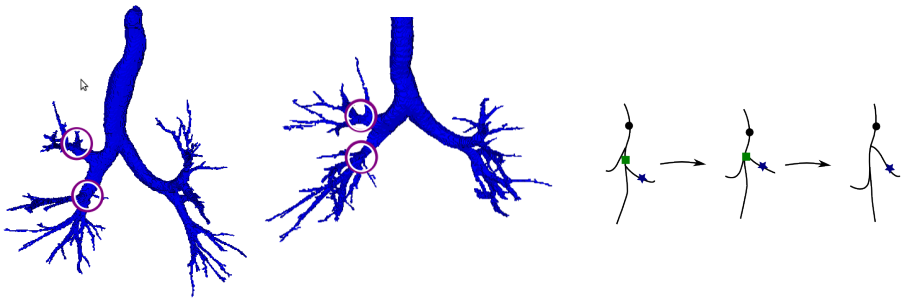
Keywords: Airway branch registration, tree-shape model, airway shape model, tree metric, tree matching

1 Introduction

Medical imaging is an important diagnostic tool, and along with this tool comes a need for automatic analysis of medical images. Tree-structures are important in this context due to their roles as delivery systems for fluids and gases, which ties them directly and indirectly to a number of diseases. For instance, chronic obstructive pulmonary disease (COPD) is tied to properties of the airway, such as the airway wall thickness [11, 14].

In order to monitor progression of diseases and determine the range of normal variation in healthy anatomical trees, we need to compare measures between scans with varying characteristics. For instance, one needs to be able to compare measurements of airway dimensions, made at specified sites in the airway,

* This work is partially funded by the Lundbeck Foundation, the Danish Strategic Research Council (PSVT projects 09-065145 and 09-061346) and by AstraZeneca AB.



(a) (b) (c)

Fig. 1. ((a-b) Examples of segmented airways (CASE31 and CASE32 in the EXACT'09 dataset). The topological structures of the two trees are different, especially the circled "trifurcation-like" regions in the (image) left hand side, shown as trifurcations in fig. 3(a). c) By tracing points (illustrated by the circle, square and star) through the geodesic deformation between two tree-shapes, we obtain a registration of points and branches of the two endpoint tree-shapes. Branches which are collapsed during the tree deformation, are not traced further, as illustrated by the square.

between several patients. In this paper, we provide a robust way to make such comparisons for airway trees by giving anatomical labels to branches in the airway based on the shape of the airway centerline tree. The technique uses a model of airway tree-shape which incorporates two notions: tree topology (i.e., parent/child connectivity) and geometry, defined by branch shapes (e.g., through landmark points).

Branch matching is, indeed, both a tree-topological and a geometric problem. Topological differences are particularly problematic for comparison and matching of anatomical trees such as the airways. These differences can come from noise in terms of spurious or missing branches, due to problems in the image recording and processing procedures. Topological differences can also come from anatomical variation between different patients, see figs. 1(a) and 1(b). Most current anatomical tree labeling methods focus, however, *either* on branch geometry *or* on tree topology. In this article we perform matching based on geodesics in an airway tree-space based on the tree-shape framework developed by Feragen et al. [3], where a varying tree topology becomes an integral part of the shape geometry. Not only is this framework geometrically very natural; it also handles tree-topological differences in a continuous, morphological way.

The airway tree model proposed in this paper consists of a shape space construction for treelike shapes, endowed with a geodesic metric. Any two treelike shapes are connected by a shortest possible deformation, or tree-space path (geodesic), whose length defines a distance between the shapes. Throughout the deformation, the initial tree changes its tree-topological structure to obtain an optimal match with the second tree. This makes the tree-shape framework well suited for branch registration in both inter- and intra-patient pairs of airways. The deformation, which is unique for sufficiently local data, induces a matching

of points along the tree-shape branches as illustrated in fig. 1(c). In a leave-one-out fashion, we generate automatic branch labels on any given airway tree by matching it with the other airway trees, which have been manually labeled by an expert. The geodesic matching is combined with a branchwise vote among anatomical labels induced by the matches, giving a robust automatic branch labeling.

The main contributions of this paper are i) the adaptation of the tree-shape framework of [3] to airway trees, giving a new version of the shape space; ii) a thorough explanation of the underlying geometric ideas, making the tree-shape model available for a broader community; iii) turning the computed geodesics into an actual branch matching; and iv) the fusion of several branch matchings through a voting scheme in order to obtain branch labelings.

The airway shape model has potential for applications beyond branch matching and labeling. Labels or geodesic distances could be used to classify airways into shape- and structure-dependent phenotypes, and the shape space framework also opens for defining new biomarkers based on the whole airway shape – topology and geometry combined. Moreover, the shape space framework used here is very general, and can be transferred to other types of data, e.g., vascular trees or medial axes, with little effort.

1.1 Related work.

Registration, branch matching and branch labeling in anatomical trees have been studied in various ways for the past decade. Some of the most successful methods are based on using association graphs [5,9,12]. Given two initial trees or graphs, their association graph is a larger graph, which contains information from both initial graphs. Branch matchings are induced by maximal cliques in the association graph. The association graph and its maximal clique are predominantly combinatorial constructions, although they can depend on geometric properties of the initial trees. Separating the analysis of geometric and combinatorial properties like this is somewhat artificial, as the geometric and topological structures play together in defining the efficiency of an anatomical tree as a space-filling structure [8,15]. Moreover, finding the maximal clique is NP hard, making exact computation intractable.

A more basic labeling method is given by van Ginneken et al. [13], where an airway tree branch labeling is made recursively, starting at the trachea, as part of the segmentation process. Labels are assigned using measures such as radius, orientation and bifurcation angle. Such a labeling approach is likely to be vulnerable to differences in topological structure, as is also noted by the authors. A different approach is that of Kaftan et al. [4], who match tree *paths* rather than branches. They avoid the difficulty with different tree-topological structures, but also lose all information stored in the topological structure. Their model does not seem to have applications beyond matching, and in particular does not generate branch labels. Smeets et al. [10] match branches from lung vessel trees using pairwise distances between nodes both in 3D Euclidean space and along the tree to generate distance matrix "fingerprints", which are matched

in order to generate a matching. Bülow et al. [2] match airway tree branches without connectivity information, using only branch shape.

These methods all focus on one out of two properties of a tree-shape: tree topology [5, 9, 12], or branch-wise geometry [2, 4, 10]. However, the airway tree is *both* topology and branch geometry. This duality is precisely what makes matching and labeling difficult. Our airway tree-shape model is ideal for modeling airway trees because it considers topology and geometry simultaneously.

The rest of the paper is organized as follows: The airway tree-shape model and the tree-shape metric are presented in sec. 2; experimental details and results are presented in sec. 3, which are discussed in sec. 4 and concluded in sec. 5.

2 The tree-shape space

The proposed method uses airway centerlines as input. Each branch centerline is associated with an edge in a combinatorial airway tree, endowed with a hierarchical tree structure (parent/child) which describes branch connectivity. Shape information is represented by edgewise attributes, in this case a fixed number of equally spaced landmark points along each branch centerline (the equal spacing length varies from branch to branch).

We build a space of airway tree-shapes based on the tree-shape model defined by Feragen et al. [3]. Any tree-shape, such as an airway tree-shape, is represented as a pair (T, f) of a combinatorial tree-structure T with edgewise shape attributes f, see fig. 2. Here $T = (V, E, r)$ is a tree with vertices V, edges $E \subset V \times V$, and a root r. The shape attributes given by n landmark points per edge are specified by a function $f\colon E \to \mathbb{R}^{3n}$.

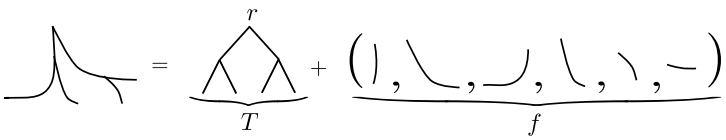


Fig. 2. A tree-shape is represented mathematically as a pair (T, f) where T is a binary, combinatorial tree and $f : E \to \mathbb{R}^{3n}$ assigns a shape-descriptor from $\mathbb{R}^{3n}$ to each edge.

2.1 Intuitive description of a geodesic tree-shape space

Our goal is to construct a continuous space of deformable trees, and this goal poses some constraints on the possible geometric structure of the space of tree-like shapes. To see this, first consider airway trees with a single fixed topological structure $T = (V, E, r)$. All such trees are described by a point in the same Euclidean product space $\prod_E \mathbb{R}^{3n}$, where each edge $e \in E$ has an associated attribute vector in $\mathbb{R}^{3n}$ describing its shape. Next, divide the set of all airway trees into classes with fixed topological structure. Each airway tree class lives

within its own Euclidean space $\prod_{E_i} \mathbb{R}^{3n}$, and the space of all airway tree-shapes is a disjoint union $X = \bigsqcup \left(\prod_{E_i} \mathbb{R}^{3n} \right)$ of shape spaces, one for each topological structure. We consider each $\prod_{E_i} \mathbb{R}^{3n}$ as a component of the larger space of all airway tree-shapes. We should tie these components together into a large shape space that connects all airway trees. In order to understand how to tie the components together, we consider how the tree-topological structure changes throughout deformations.

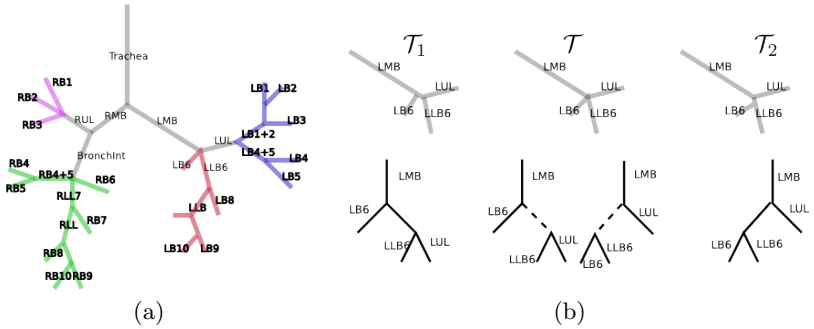


Fig. 3. (a) This figure is best viewed in color. In the first generations of the airways, the branching structure is fairly pre-determined due to anatomy, and the branches have names. (b) The tree-shape $\mathcal{T}$ can be obtained as a limit of a sequence where a branch is disappearing – for instance, $\mathcal{T}$ can be obtained by shrinking a branch in either $\mathcal{T}_1$ or $\mathcal{T}_2$, or a number of other tree-shapes $\mathcal{T}_i$. This illustrates how trifurcations can be interpreted as pairs of bifurcations; their combinatorial representations are shown in the bottom row. This causes some problems in the representation space X defined in sec. 2.2, since, for instance, the path from $\mathcal{T}_1$ to $\mathcal{T}$ to $\mathcal{T}_2$ is impossible in X. Passing from $\mathcal{T}_1$ or $\mathcal{T}_2$ to $\mathcal{T}$ is easy, but the two representations of $\mathcal{T}$ correspond to different points in X, and the path $\mathcal{T}_1 \to \mathcal{T} \to \mathcal{T}_2$ is not possible in X. This path is, however, possible in the quotient space $\bar{X}$, where different representations of the same tree are glued together in one point. We define $\bar{X}$ to be the space of treelike shapes.

For example, in fig. 3(a), the LMB branch is shown as the parent of a trifurcation, with the LUL, LLB6 and LB6 branches emanating from it. In reality, one will not find a trifurcation, but a pair of bifurcations, e.g., as seen in the two airway subtrees $\mathcal{T}_1$ and $\mathcal{T}_2$ in fig. 3(b). A geodesic deformation from $\mathcal{T}_1$ to $\mathcal{T}_2$ should interchange the order of the LUL and LB6 branches by passing through a tree of the type $\mathcal{T}$ shown in fig. 3(b). This tree can be obtained as a limit of a sequence of trees with a fixed structure. For instance, such a sequence can start at $\mathcal{T}_1$ or $\mathcal{T}_2$ and successively shrink the small branch until collapsing onto one of the representations of $\mathcal{T}$. Similar sequences can be found for many different topological structures, represented by tree-shapes $\mathcal{T}_i$, where slowly deleting certain edges in $\mathcal{T}_i$ converges towards a tree of type $\mathcal{T}$. By following sequences within different components of the tree-space, where all the tree-shapes in the i^{th}

component have the same structure as $\mathcal{T}_i$, we reach the same tree-shape $\mathcal{T}$. This must mean that in order to have a tree-space where the trees can continuously deform from one topological structure to another, as in fig. 3(b), the tree-space components of the $\mathcal{T}_i$ must all intersect along the component of $\mathcal{T}$. In the next section, we shall see how such a tree-space can be defined mathematically.

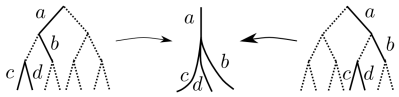


Fig. 4. Trees which are *not* binary, or which are smaller than T, are represented by cancelling extra edges, endowing them with the attribute zero (dotted lines). This leads to several representations of the same tree-shape. By identifying those different representations that represent the same shape, we construct our shape space as the quotient space $\bar{X}$.

2.2 Mathematical definition of the tree-shape space

Having established an intuitive understanding of the geometry of the tree-space, we shall give a more concrete definition. Recall that any tree-shape is represented by a pair (T, f) consisting of a combinatorial tree T and edge shape attributes f. In order to study trees of different sizes and topologies in one unified setting, all shapes are represented by a fixed combinatorial tree $T = (V, E, r)$, which is *large enough*; e.g., if all the tree-shapes have depth N, then T could be the full binary tree of depth N. Smaller trees are represented in the tree-space by endowing extra edges with zero attributes, see fig. 3(b). All trees are represented by a point in the Euclidean representation space

$$X = \prod_{e \in E} \mathbb{R}^{3n}. \tag{1}$$

Assume, moreover, that T is binary. All non-binary tree-shapes can be represented using a binary combinatorial tree by collapsing internal branches, as described in fig. 4. The space $X = \prod_{e \in E} \mathbb{R}^{3n}$ contains every tree-shape represented at least once. Some tree-shapes are even represented at several points in X, and as a consequence, some natural tree-shape deformations cannot be represented as paths in X, see fig. 3(b). This problem is solved by generating an equivalence relation $\sim$ on X where different representations of the same shape are identified; now the space of tree-shapes is defined as the quotient space $\bar{X} = X/\sim$.

Definition 1. *We say that two representations in X define the "same tree-shape" when the following holds: Starting with the two tree-shape representatives $x_1 = (T, f_1)$ and $x_2 = (T, f_2)$, remove all branches with zero attribute, and consider the resulting (possibly no longer binary) tree-shape representations $(T_1, \tilde{f}_1)$ and $(T_2, \tilde{f}_2)$. Here, T_1 and T_2 are ordered, combinatorial, rooted trees. If, up to*

a topology-preserving reordering of the branches, these two attributed trees are exactly the same, then the tree-shapes are the same and the two representations are defined to be equivalent. *We write* $x_1 \sim x_2$.

All equivalent points are identified by forming the *quotient space* $\bar{X}$ of the equivalence relation [1], that is:

$$\bar{X} = X/\sim = \{\bar{x} | x \in X\}, \tag{2}$$

where

$$\bar{x} = \{z \in X | z \sim x\}. \tag{3}$$

The points in $\bar{X}$ are equivalence classes $\bar{x}$ in X – by abuse of notation we use $\bar{x}$ to denote both points in $\bar{X}$ and subsets of X. Geometrically, this construction corresponds to gluing together those points in the representation space X that represent the same tree-shape. As most of the points only have one representative, this creates a new space $\bar{X}$ with self intersections at points with several representatives. Many geometric properties of X are inherited by $\bar{X}$. The quotient space $\bar{X}$ is the space of treelike shapes.

2.3 Metrics on the shape space

The Euclidean metric on X induces a standard *quotient metric* d on $\bar{X}$ [1], called Quotient Euclidean Distance (QED), defined as follows:

$$d(\bar{x}, \bar{y}) = \inf_{k \in \mathbb{N}} \left\{ \sum_{i=1}^{k} \|a_i - b_i\| : a_1 \in \bar{x}, b_i \text{ equivalent to } a_{i+1}, b_k \in \bar{y} \right\}. \tag{4}$$

Here, the norm $\|\cdot\|$ denotes the Euclidean norm on X, and $\bar{x}$ is the equivalence class of the point $x \in X$. Thus, when we write $a_1 \in \bar{x}$, this means that a_1 is a point belonging to the equivalence class $\bar{x}$ as a subset of X, or equivalently, $a_1 \sim x$.

One interpretation of eq. 4 is that a QED geodesic consists of a sequence of k Euclidean lines, which are cut and concatenated whenever the geodesic deformation switches between two representations b_i and a_{i+1} of the same tree. Typically, these identified points correspond to internal topological transitions in the tree-shape structure. The infimum is taken over all possible concatenations of lines for any k. E.g., the geodesic from $\mathcal{T}_1$ to $\mathcal{T}_2$ in fig. 3(b) consists of the Euclidean line from $\mathcal{T}_1$ to the first representative of $\mathcal{T}$, concatenated with the Euclidean line from the second representative of $\mathcal{T}$, to $\mathcal{T}_2$.

The QED metric is locally very well-behaved. In particular, geodesics between data points and various forms of average shapes will exist and be unique [3]. This makes the tree-shape space and its geodesic deformations well suited for registration. The tree-space construction described above is completely general, and applies to any trees with continuous edge attributes, not only shapes.

2.4 Airway tree-shape space

In order to compute the geodesic connecting two $3D$ tree-shapes such as airways, we consider all possible branch orders on the trees in order to find an optimal branch alignment. The number of possible orders grows exponentially with the size of the trees, resulting in computational difficulties. In the case of airway trees, however, some branches are easy to identify, e.g., the main bronchi and some of the lobar bronchi seen in fig. 3(a). We significantly reduce computational complexity by identifying and fixing the branches $E_{\text{fixed}} = \{\text{RMB, LMB, RUL, BronchInt, LB6, LLB6, LUL}\}$, which are present and easy to identify in most data trees. In this way, we consider topological variation only in the lobar subtrees T_{RUL}, T_{LUL}, $T_{\text{BronchInt}}$, T_{LLB6}, T_{LB6} following these, shown in colors in fig. 3(a). The airway shape space is then

$$\left(\prod_{e\in E_{\text{fixed}}} \mathbb{R}^{3n}\right) \oplus \overline{\left(\prod_{e\in E_{\text{RUL}}} \mathbb{R}^{3n}\right)} \oplus \cdots \oplus \overline{\left(\prod_{e\in E_{\text{LLB6}}} \mathbb{R}^{3n},\right)} \tag{5}$$

where the first component is Euclidean and the others are quotient spaces. The symbol $\oplus$ denotes direct sum, or Cartesian product.

2.5 Computing geodesics

Computing the geodesics from eq. 4 is generally NP hard. The number k of Euclidean concatenations that need to be checked will, in practice, be bounded for any given pair of trees, but it will grow exponentially with the number of edges in the trees. We make an approximation by bounding k in eq. 4 for each lobar subtree. That is, we fix some number $K \in \mathbb{N}$, and approximate the distance in eq. 4 by computing

$$d(\bar{x},\bar{y}) = \inf_{k\leq K}\left\{\sum_{i=1}^{k} \|a_i - b_i\| : a_1 \in \bar{x}, b_i \text{ equivalent to } a_{i+1}, b_k \in \bar{y}\right\}. \tag{6}$$

This corresponds to assuming that the tree-shape represented throughout geodesic will undergo at most K internal topological changes. In order to compute the approximated distance, a naive implementation is to list all the allowed combinations of topological changes, to compute the shortest version of a path going through each of those particular changes, and choose the shortest among the resulting paths. This is equivalent to Algorithm 1 from [3].

2.6 Branch label extraction

Using a geodesic deformation from a labeled airway tree to an unlabeled one, as seen in fig. 1(c), labels from a labeled tree can be propagated to the branches of an unlabeled tree.

For data with large variation in topology, we would normally need to choose a rather high bound K on k in eq. 6, in order to obtain the true geodesics between data points. This is punished by a large increase in computation time. We avoid

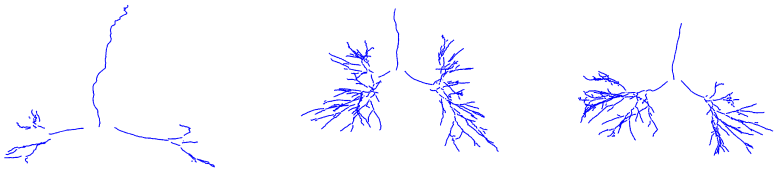


Fig. 5. The centerline trees of CASE26, CASE34 and CASE39, respectively. Note the large variation in size and topological structure; in particular CASE39 is missing the upper lobar trees.

this problem by matching with multiple labeled trees, and thus extracting a number of candidate labels for each branch. This is followed by a voting scheme, where nearby trees, for which the approximation is good, are expected to win the vote. Given a training set of airway trees with manually assigned labels, a new airway tree is first matched with all the trees in the training set, and sets of labels are propagated to the branches of the unlabeled tree. Each branch is assigned a pool of labels, among which the majority vote is selected as a new, automatically extracted branch label.

3 Experimental validation

Experiments were made on airways from the EXACT'09 challenge test set [7], consisting of 20 CT scans from 15 different subjects, labeled as CASE21 - CASE40. The EXACT'09 challenge was a segmentation competition, and the CT scans come from a wide range of sources, states of breathing, and are processed using different methods [7]. As a result, the segmented airway trees display great variation in size, shape and noise level.

The airway trees were segmented from the CT scans using a voxel classification based airway tree segmentation algorithm by Lo et al. [6]. The centerlines were extracted from the segmented airway trees using a modified fast marching algorithm, directly giving a tree structure. Due to the segmentation method, the centerline trees are not connected, but have gaps at each bifurcation, as seen in fig. 5. This is not a problem, as the proposed matching method relies on the edge shapes, which are sufficiently well described in the disconnected model, along with parent/child connectivity information.

Leaves with segmented volume less than 10 mm^3 were assumed to be noise and pruned away, and the centerlines were sampled with 6 equidistant landmark points along each edge. For each edge, the landmark points were translated so that the first landmark point was aligned with the origin. In this way, large differences in edges of low generation do not affect the whole subtree following them. To account for variation in size, each airway tree was normalized by the constant scaling factor 1/length(LMB), chosen since the LMB branch is present

CASE	21	22	23	24	25	26	27	28	29	30
% correct	75	88.2	92.9	80	77.8	86.7	88.9	94.4	66.7	89.5
# correct	12	15	13	12	14	13	16	17	14	17
CASE	31	32	33	34	35	36	37	38	39	40
% correct	90	76.5	88.9	100	83.3	78.9	66.7	80	30	76.5
# correct	18	13	16	13	15	15	12	8	4	13

Table 1. Results from the branch labeling; the percentage of correctly labeled branches among all branches labeled by the algorithm, and the number of branches correctly labeled by the algorithm. CASE39 is an outlier: both upper lobes were missing from the segmentation, making the algorithm fail. When the outlier is left out, we correctly label 83% of the branches on average.

and easy to measure in all segmentations. The trachea was left out in the experiment due to varying cut-off points. The next few branches (RMB, LMB, RUL, BronchInt, LB6, LLB6, LUL) were detected based on the orientation and extent of their subtrees. Branches from the first $6-7$ generations were matched, whenever they were present in the segmentation.

3.1 Branch labeling

Each airway tree was aligned with all other airway trees in the dataset through computing geodesics with $K = 2$ for each lobar subtree (that is, permitting one structural transition in each of the five lobar subtrees). When computing the distances, not only the length of the geodesic from one tree to another, but also the geodesic from the first tree to the second were recorded. Using the geodesic we obtained a branch-wise matching by recording which edges are mapped where.

The reference airway trees from EXACT'09 were labeled manually by a trained image analyst. Up to 34 labels were assigned according to a standard nomenclature used in bronchoscopy. All cases were reviewed by a pulmonologist after labeling. For each pair of airway trees, a ground truth branch matching was induced from the anatomical labels. Several branches found by the registration were not labeled by the human experts, and were left out in evaluation.

The voting scheme was made based on the 20-airway dataset in a leave-one-out fashion, where each airway tree was matched with the remaining 19 and given a branch labeling based on the majority vote. In this way, labels down to the sixth generation of each airway tree were computed using the rest of the airway trees as training set. The best label for any given branch was obtained by voting among the labels. Branch labels with less than 55% voting consensus or less than 4 votes in total were discarded. The results are found in table 1; the overall success rate is 83%. We can obtain better scores by insisting on a higher majority vote; however, this implies labeling fewer branches.

4 Discussion

Tschirren et al. [12], van Ginneken et al. [13] and Bülow et al. [2] performed labeling on airway trees with success rates of 97%, 90% and 69%/40% (using

different features), respectively. Our success rate is 83%, which – taking the level of variation in the dataset into account – is high. Since the datasets used are all different, a direct comparison of percentages is not possible.

The large variance in our results is natural since the EXACT'09 data come from a wide range of sources, and are made with different subjects, scanners, scan- and reconstruction protocols etc, as described in [7, table 1]. This poses a great challenge: the structural differences between the segmented airway trees are very large, as illustrated in figs. 5 and 6, and the amount of noise is very different from tree to tree. We might have dealt with this problem by making a better approximation of the QED metric. Instead, based on the hypothesis that trees which are close together will be topologically similar, we choose to use a coarse but efficient approximation combined with a voting scheme.

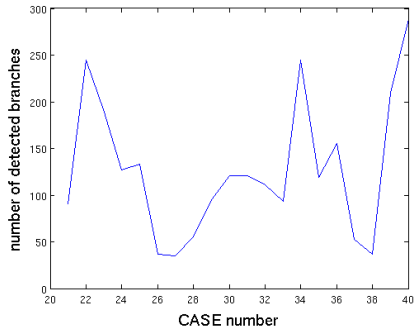


Fig. 6. The number of detected branches in the EXACT'09 airway trees is very variable. Note also that this number is affected by the level of structural noise, and might not directly correspond to the number of anatomical branches.

5 Conclusion

Based on a geometric model for shapes with a treelike structure, we have developed a technique for automatic inter- and intra-patient registration of airway tree centerlines. The method has been evaluated by performing an airway branch labeling on the EXACT'09 dataset. In spite of the variation in the dataset, our labeling results are good, illustrating the potential of the shape framework.

The shape space framework is very general, and the tree-space can be enriched with additional attributes if wanted. There are very few airway and vessel tree labeling algorithms available, most of which are ad hoc, specialized heuristics. In contrast, our proposed method takes a principled approach, which is easy to generalize to other biological tree-structures, e.g.,vascular structures. This makes the proposed method both novel and important.

The airway shape model has potential for many applications beyond branch matching and labeling. The shape space has good properties for statistical analysis such as computation of means and modes of variation. Geodesic distances

could be used to classify airways into shape- and structure-dependent phenotypes, and the shape space framework also opens for defining new imaging biomarkers based on the whole airway shape – topology and geometry combined. These extensions are, however, by no means straightforward, as we are working in the non-smooth domain of tree-shape space. Further development of the shape space, numerical methods for tree-shape computations, and their applications are all topics of future work.

References

1. M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Springer-Verlag, 1999.
2. T. Bülow, C. Lorenz, R. Wiemker, and J. Honko. Point based methods for automatic bronchial tree matching and labeling. In *SPIE Med. Im.*, volume 6143, pages 225–234, 2006.
3. A. Feragen, F. Lauze, P. Lo, M. de Bruijne, and M. Nielsen. Geometries in spaces of treelike shapes. In *ACCV*, pages 671–684, 2010.
4. J.N. Kaftan, A.P. Kiraly, D.P. Naidich, and C.L. Novak. A novel multi-purpose tree and path matching algorithm with application to airway trees. In *SPIE Med. Im.*, volume 6143, pages 215–224, 2006.
5. H. Kitaoka, Y. Park, J. Tschirren, J. M. Reinhardt, M. Sonka, G. McLennan, and E. A. Hoffman. Automated nomenclature labeling of the bronchial tree in 3D-CT lung images. In *MICCAI (2)*, pages 1–11, 2002.
6. P. Lo, J. Sporring, H. Ashraf, J.J.H. Pedersen, and M. de Bruijne. Vessel-guided airway tree segmentation: A voxel classification approach. *Medical Image Analysis*, 14(4):527–538, 2010.
7. P. Lo, B. van Ginneken, J.M. Reinhardt, and M. de Bruijne. Extraction of Airways from CT (EXACT09). In *The 2nd Intl WS Pulm. Im. Anal.*, pages 175–189, 2009.
8. B. Mauroy, M. Filoche, E. R. Weibel, and B. Sapoval. An optimal bronchial tree may be dangerous. *Nature*, 427:633–636, 2004.
9. J. H. Metzen, T. Kröger, A. Schenk, S. Zidowitz, H-O. Peitgen, and X. Jiang. Matching of anatomical tree structures for registration of medical images. *Im. Vis. Comp.*, 27:923–933, 2009.
10. D. Smeets, P. Bruyninckx, J. Keustermans, D. Vandermeulen, and P. Suetens. Robust matching of 3d lung vessel trees. In *MICCAI 2010, Proc 3rd Intl WS Pulm. Im. Anal.*, pages 61–70, 2010.
11. L. Sørensen, P. Lo, A. Dirksen, J. Petersen, and M. de Bruijne. Dissimilarity-based classification of anatomical tree structures. In *Information Processing in Medical Imaging*, LNCS, 2011.
12. J. Tschirren, G. McLennan, K. Palágyi, E. A. Hoffman, and M. Sonka. Matching and anatomical labeling of human airway tree. *IEEE TMI*, 24(12):1540–1547, 2005.
13. B. van Ginneken, W. Baggerman, and E. van Rikxoort. Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In *MICCAI 2008*, pages 219–226, 2008.
14. G. R. Washko, T. Dransfield, R. S. J. Estepar, A. Diaz, S. Matsuoka, T. Yamashiro, H. Hatabu, E. K. Silverman, W. C. Bailey, and J. J. Reilly. Airway wall attenuation: a biomarker of airway disease in subjects with COPD. *J Appl Physiol*, 107(1):185–191, 2009.
15. E. R. Weibel. What makes a good lung? *Swiss Med. Weekly*, (139(27-28)):375–386, 2009.

Deformable Organisms and Error Learning for Brain Segmentation

Gautam Prasad[1,2], Anand A. Joshi[3],
Albert Feng[1], Marina Barysheva[1] Katie L. McMahon[4],
Greig I. de Zubicaray[5], Nicholas G. Martin[4], Margaret J. Wright[4],
Arthur W. Toga[1], Demetri Terzopoulos[2], and Paul M. Thompson[1]

[1]Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA, USA
[2]Department of Computer Science, UCLA, Los Angeles, CA, USA
[3]Signal and Image Processing Institute, University of Southern California, Los Angeles, CA, USA
[4]Queensland Institute of Medical Research, Brisbane, Australia
[5]School of Psychology, University of Queensland, Brisbane, Australia

Abstract. Segmentation methods for medical images may not generalize well to different data sets or tasks, hampering their utility. We attempt to remedy these issues using deformable organisms to create an easily customizable segmentation plan. This plan is developed by borrowing ideas from artificial life to govern a set of deformable models that use control processes such as sensing, proactive planning, reactive behavior, and knowledge representation to segment an image. The image may have landmarks and features specific to that dataset; these may be easily incorporated into the plan. We validate this framework by creating a plan to locate the brain in 3D magnetic resonance images of the head (skull-stripping). This is important for surgical planning, understanding how diseases affect the brain, conducting longitudinal studies, registering brain data, and creating cortical surface models. Our plan dictates how deformable organisms find features in head images and cooperatively work to segment the brain. In addition, we use a method based on Adaboost to learn and correct errors in our segmentation. We tested our method on 630 T1-weighted images from healthy young adults, evaluating results using distance and overlap error metrics based on expert gold standard segmentations. We compare our segmentations with and without the error correction step; we also compare our results to three other widely used methods: BSE, BET, and the Hybrid Watershed algorithm. Our method had the least Hausdorff distance to expert segmentations on this dataset, but included slightly more non-brain voxels (false positives). Our framework captures diverse categories of information needed for skull-stripping, and produces competitive segmentations.

Keywords: deformable organisms, segmentation, MRI, Adaboost, Hausdorff, overlap, registration

1 Introduction

Deformable organisms label objects in images by integrating high level control mechanisms into a segmentation plan. They combine sensing, knowledge representation, reactive behavior, and proactive planning so our devised organisms may cooperatively segment an image. Deformable organisms were introduced into medical imaging by [13] who combined ideas from artificial life [29] and deformable models [14, 30]. Since their introduction, deformable organisms have been used for limb delineation [18], and segmentation of the spinal cord [16], vasculature [17], and corpus callosum in the brain [6]. [15] created a deformable organisms framework using the Insight Toolkit (ITK) [8], but we did not use it here, as its capabilities were too basic for our application. In contrast to several brain segmentation methods that work with low-level image processing and computer vision techniques, our deformable organisms can incorporate high-level knowledge and expectations regarding image data. In addition, almost every step of the plan presented here differs significantly from the one we previously presented [21].

Segmenting brain from non-brain tissues (such as the eyes, skull, scalp, and neck) in magnetic resonance (MR) images of the head is a vital pre-processing step for many types of image analysis. Accurate masks of the brain are helpful for longitudinal studies [22], for multi-subject analyses of brain structure and function [31], and as a pre-processing step prior to cortical surface modeling [32], surgical planning [4], and brain registration [34].

The process of segmenting brain versus non-brain tissue on MRI is commonly referred to as "skull-stripping" (although, strictly speaking, the skull generates almost no signal on T1-weighted MRI and the scalp and meninges are the main tissues removed). This has traditionally been done manually by trained experts, or by automated methods that are subsequently corrected by hand. Manually-created masks may also be used as gold standard delineations to validate performance of skull-stripping methods based on different principles. Though many approaches have been developed for this task, time consuming manual clean-up of these generated masks is almost always required. Most published methods do not perform well on all datasets, making improvements over existing methods critical.

There are a variety of existing skull-stripping methods. The Brain Extraction Tool (BET) [28] evolves a deformable model to find the boundary of the brain. It provides a robust way to find the boundary in unclear regions but does not incorporate prior knowledge of the brain's shape. The Brain Surface Extractor (BSE) [25] uses edge detection and morphological operations to find the brain/non-brain boundary. BSE quickly extracts the brain from an image but may include extra material in the mask, as it sometimes fails to remove connections between the brain and surrounding tissue. The Hybrid Watershed Algorithm (HWA) [24] uses the watershed algorithm to find the brain region, then fits a deformable model to the region, and finally deforms it based on a statistical atlas and geometric constraints. These methods have also been analyzed

in [1]. We chose these methods as they are the most widely used and are part of larger neuroimaging toolkits.

We create a deformable organism plan that governs a collection of organisms to segment different parts of the head and brain. The organisms evolve dynamically in the images and cooperatively compute an accurate and robust segmentation of the brain. We then use a learning method, based on an Adaboost wrapper, [33], to classify the error in our method. We evaluate the effectiveness of this additional error correction step in improving our segmentations. We test our method with 630 T1-weighted MR images from healthy young adults, aged 20-30 years. We compare our approach to three widely used methods and we validate our results using distance, overlap, and error metrics.

2 Methods

Our deformable organisms method aims to segment the brain in T1-weighted MR images of the head. We describe our deformable organism plan to segment the brain, a way to learn and correct errors in our method, validation metrics to compare our results to the gold standard and to other widely-used methods, and our experimental results.

2.1 Deformable Organisms

Deformable organisms are organized in five different layers that combine control mechanisms and different representations to segment an image. We adapt this general approach for segmenting the brain.

Geometry and Physics We represent each organism as a 3D triangulated mesh. These meshes are initialized on a standard brain template image. Our template was selected from the 40 images in the LONI Probabilistic Brain Atlas (LPBA40) [26], which have corresponding manual segmentations for 56 structures, and have manual delineations of the brain boundary. In the image we selected from this set, the voxels lying in each of our regions of interest are labeled. We fit our organisms to these labels to create a mesh using a marching cubes method [12] that goes through the image. The mesh is made up of polygons representing the border of the regions, which are then fused together.

These meshes deform to fit the 3D region that their corresponding organism is modeling. This iterative process moves each of the mesh's vertices along its normal direction with respect to the mesh surface. The surface is smoothed at every iteration using curvature weighted smoothing [2, 19]. This smoothing technique attenuates noise through a diffusion process as

$$\frac{\partial \mathbf{S}}{\partial t} = \lambda L(\mathbf{S}), \tag{1}$$

where S is the mesh, surface, or manifold and L is the Laplacian, which is equivalent to the total curvature of the surface. This Laplacian is linearly approximated

by

$$L(\mathbf{v}_i) = w_{ji} \sum \mathbf{v}_j - \mathbf{v}_{i_{j \in N(i)}}, \tag{2}$$

where v_i is the vertex i in the mesh, $N(i)$ are the neighbors of i, and w_{ji} is a weight proportional to the curvature between vertices j and i. We smooth the mesh to constrain its deformations to prevent intersections and artifacts from corrupting the boundary. On the boundary, we sample from the surrounding image.

Our prior based on an image from LPBA40 precludes us from using the Segmentation Validation Engine tool [27] to validate our method as it would lead to an unfair bias that would be favorable to our method; instead we use other metrics (below).

Perception Our organisms "sense" the encompassing image by sampling its intensities at vertices of the mesh. The vertices composing the mesh have real-valued coordinates, so we used nearest neighbor interpolation to find the intensities that correspond to them in the discrete grid of voxels in the image. The images may be any of the subject-derived volumes, which include the threshold image, 2-means classified image, 3-means classified image, or gradient image.

Paramount to the perception layer is the organisms' ability to sense each other's locations. We locate the voxels an organism resides on using a 3D rasterization algorithm [20, 5] that efficiently computes these values. These locations allow our organisms to dynamically change the way they deform based on their own positions and intensities of the original T1-weighted subject image.

Motor Control We move the vertices of the mesh along their normal direction with respect to the mesh surface by analyzing a set of intensities along this normal line. We describe the evolution of our mesh or surface $S(i,t)$ with respect to time t, where i is a vertex or point on the surface, as

$$\frac{\partial \mathbf{S}(i,t)}{\partial t} = F(\mathbf{P}, \mathbf{n}, \mathbf{I}_d, b(c,l)_j)\mathbf{n}, \tag{3}$$

with F being the speed of evolution. F samples a set of positions P along point i's normal and interpolates these values from any of the derived images I_d, where d specifies the set of derived image. That set consists of the threshold image (t), 2-means image (2), 3-means image (3), and gradient image (g). The function $b(c)_j$ specifies any of a number of behaviors and decides how to move the point i on the surface by analyzing the sampled intensities l subject to a set of constraints, and weights the movement by the scalar c.

In practice, we evolve each vertex by the amount specified by F along its normal and progress through time by iterating through all vertices in the mesh until there is no longer any significant movement.

Behavior Our behaviors are a higher level of abstraction to indicate how the organisms function and what information they need to find. Behaviors may prescribe a function for organisms to be attracted to or repelled from landmarks or help converge on the boundary of an object. The functions for our behaviors had specific tasks in mind in the context of skull-stripping but are general and simple enough for repeated use in any segmentation task.

We developed six behaviors that play an important role in almost every step of our skull-stripping plan.

1. We create a behavior that analyzes a binary image and locates a boundary in these images. It contracts if a vertex corresponds to an off value and expands if it corresponds to an on value, and may be described as

$$b(c,l)_1 = \begin{cases} -c & \text{if } l_i = 0 \\ c & \text{otherwise} \end{cases} \tag{4}$$

   In this case the set l consists of a single value l_i, the value of the binary image that corresponds to the vertex i.
2. Our second behavior moves a vertex outwards if its corresponding intensity value is q and may be described as

$$b(c,l)_2 = \begin{cases} c & \text{if } l_i = q \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

   Its purpose is to expand into an area of an image with voxels having intensities q. In addition q may be a set of labels that are appropriate for expansion.
3. Our third behavior is customized to move our skin mesh through the skull and skin into the cerebrospinal fluid (CSF). It contracts itself further if the boundary intersects another organism (in our plan, the other organisms are the eyes and we wish to deform through those areas) and will check the intensities (l) along the normal for those that correspond to CSF. If CSF markers m are found (signified by very low intensity values and a specific label in our k-means images) we contract the mesh, and if they are deficient, we expand the mesh. This behavior may be represented as

$$b(c,l)_3 = \begin{cases} -c & \text{if } l_i \text{ intersects other organism} \\ -c & \text{if } m \in l \\ c & \text{if } m \notin l \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

   In our framework, the intensities, l, are sampled from the k-means images and m is the label corresponding to CSF. The sampled points are locations inside the surface with respect to i.
4. the fourth behavior we created was designed for the brain organism to locate the boundary of the brain. It contracts vertices away from other organisms, contracts if there are CSF marker values (m) present within the surface, expands if the label value at i is not q, and expands if the gradient intensity

at l_i is greater than or equal to a threshold t. The precedence of these constraints is ordered as

$$b(c,l)_4 = \begin{cases} -c & \text{if } l_i \text{ intersects other organism} \\ -c & \text{if } m \in l \\ c & \text{if } l_i \neq q \\ c & \text{if } l_i \leq t \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The values at l are sampled from either the 3D rasterization image of the other organisms' meshes, the 3-means classified image, or the gradient image.

5. Our final behavior is the same as behavior 2 but instead of expanding it contracts.

$$b(c,l)_5 = \begin{cases} -c & \text{if } l_i = q \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Cognition We create a plan of different behaviors to perform a segmentation task. The plan may dynamically activate different behaviors depending on what features the organisms were able to find in the image. Our plan to skull-strip the brain is one such plan.

2.2 Skull-Stripping Plan

Our skull-stripping plan combines our image processing and deformable organisms to create objectives in the image to extract the locations of different regions, culminating in extracting the boundary of the brain. In what follows, we describe each step in detail and how it depends on previous knowledge obtained by organisms. This is just one plan and may be fashioned for any type of segmentation or specifics of the data. Table 1 summarizes the steps each organism takes during the segmentation.

1. We begin by registering the subjects T1-weighted MR image to the template we selected from the LPBA40. This registration step is important to transform subject images into a standard coordinate space as our organisms are tuned (iterations for deformations and labels for k-means classification) to images roughly corresponding to our template. Our template incorporates prior information and may be changed by users who need something closer to their own data. It provides initial locations and shapes for our skin, eye, and brain organisms.
   We used an affine transformation for registration provided by FMRIB's Linear Image Registration Tool (FLIRT) [9]. FLIRT uses the correlation ratio [23] as the metric between the two images that takes the form

$$\frac{1}{Var(Y)} \sum \frac{n_k}{N} Var(Y_k)_k . \tag{9}$$

   Y represents one of the images, $Var(Y)$ is the variance of Y, Y_k is the k-th iso-set i.e. the set of intensities in Y at positions where the other image X has

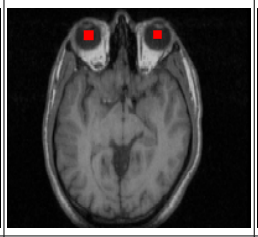
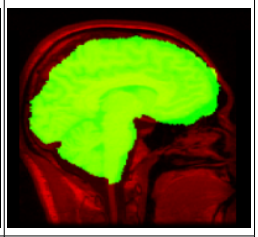
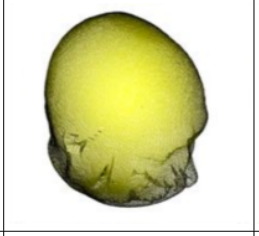
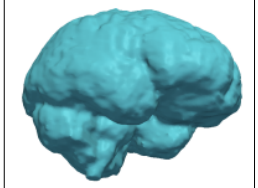
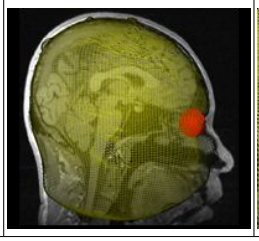
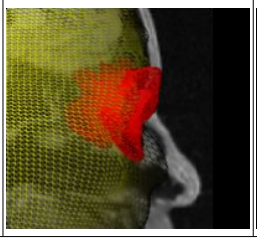
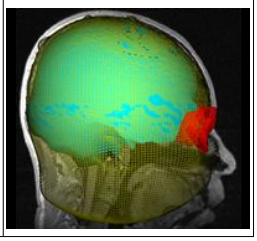
intensities in the kth bin, n_k is the number of values in Y_k with $N = \sum_k n_k$. This cost is optimized to find a 12-parameter affine transformation.

In addition, we compute the inverse transformation to take the subject image back to its native space at the end of the segmentation.

2. We find the location or boundary of the skin with the skin organism. Its initial shape is of the skin of our template image found using the marching cubes method. We dilated this substantially to ensure we encompass the head of any subject registered to the template. Our template-fitted mesh needs to be further refined to fit our subject. To do this, we analyze the subject's intensities and apply a threshold to create a binary image masking out the head. We also use behavior 1 to sense the threshold image and evolve our skin organism's mesh to find this perimeter. We iterate the deformations dictated by behavior 1 (applying smoothing at every step) until there is no significant movement of the surface or we reach a maximum iteration bound. We specify this bound based on images being reasonably aligned to the template, an approach we used for all our deformations. The adjacent eyes are handled in a similar manner.
3. Our eye organisms find the eye boundary by sensing the 3-means classified image. We initialize the eye organisms' meshes by fitting them to our template and eroding them to make sure they lie within a subject image's eyeballs. The eyeball locations are found by the organisms sensing the 3-means image with behavior 2, which chooses a label found in the eyeball region.
4. Our next step locates the cerebrospinal fluid (CSF) that surrounds the brain. We achieve this using our skin organism by contracting its mesh into the head through the skin and skull. The skin and skull locations are roughly classified in our 2-means image and we apply behavior 3 to sense it and find the CSF boundary. To further refine this boundary the behavior also makes the skin organism deform through the eye organisms because there is more information about the CSF location by the eyes.
5. This step finds the tissues surrounding the eyeballs that need to be excluded from the brain delineation. We attain this goal by expanding the eye organisms further by sensing the 2-means image along with behavior 2 again, this time behavior 2 looks for a different label in the classified image, one that gives an understanding of tissues around the eyes. The eyes now furnish a better understanding to locate the brain.
6. We complete our plan by finding the brain using our brain organism. Every step in the plan supports of this step and all of our knowledge up to this point will come into play. Our brain organism begins by sensing the 3-means image and gradient image with behavior 4. The behavior is cognizant of the other organisms' locations and uses them to constrain its evolution. With the completion of behavior 4 we further refine the boundary by sensing the 3-means image again with behavior 5, which results in the brain being encapsulated by our brain mesh. The mesh is then converted by the 3D rasterization scheme to a binary volume to which we apply our inverse

transform from FLIRT to bring the subject's delineation back into its native space to complete the segmentation.

Table 1: Plan for each organism to skull-strip an MRI image.

	Skin Organism	Eye Organism	Brain Organism
Initialization			
Deformation 1			
Deformation 2			

We make a binary volume of the brain organism and apply the inverse transformation back to the subject image space, completing the segmentation.

3 Error Learning

We are able to learn the types of errors our method makes, by comparing the masks it generated with expert manual delineations. [33] introduced an algorithm using Adaboost [3] to learn a weighting of a set of features used to classify if a voxel has been correctly classified by a prior algorithm. This 'Adaboost wrapper' algorithm uses a set of corresponding automated and manual segmentations, as well as intensity images to find features that lie in regions that the first-pass method incorrectly classifies. We use this algorithm to learn situations in which our method makes errors and thereby improve the segmentation.

4 Validation

The masks from the deformable organism method are compared with the gold standard manual segmentations using standard distance, overlap, and error met-

rics. We use the Hausdorff distance measure [7] to find the distance from the furthest point in the deformable organism method mask to the closest point of the mask in the manual delineation. We also compare the expert and automated masks using the Jaccard coefficient (union overlap), Dice coefficient (mean overlap), false negative rate, and false positive rate described in [11].

We compared the differences in metric values across methods using paired-sample t-tests to understand if their results were statistically different.

5 Experiments

We tested our deformable organism skull stripping method on our set of 630 manually-labeled subject images. In addition, we also ran BET (BET2, FSL 4.1.5, default parameters), BSE (BSE 10a, default parameters), and the Watershed algorithm (Freesurfer 5.0.0, default parameters), and assessed their errors using standard distance-based, overlap, and error metrics.

Typically patterns of error in our method were learned by selecting a subset of our segmentation results and using the error learning algorithm (the 'Adaboost wrapper' approach). We then segmented a new subset of images with the error classified and corrected. We repeated this experiment 10 times, using 10 random images from our results to train and 10 random (but non-overlapping) images to test. Masks were then compared to expert ground truth before and after error correction; note that the test set of images was independent of those used for training the error correction step.

6 Results

6.1 Subject Data

Our subject data consisted of 630 T1-weighted magnetic resonance (MR) images from healthy young adults, between 20 and 30 years of age. These images are from Australian twins, and have been used in numerous prior analyses [10].

Each of the images had been manually skull-stripped by a neuroanatomically trained expert. These manual labels were used as the gold standard to compare with automatic segmentation results of our method and the other 3 widely-used methods. The subjects were scanned with a 4-Tesla Bruker Medspec whole-body scanner. 3D T1-weighted images were acquired using a magnetization-prepared rapid gradient echo sequence, resolving the anatomy at high resolution. Acquisition parameters were: inversion time (T_I)/repetition time (T_R)/echo time (T_E)=700/1500/3.35 ms, flip angle=8°, slice thickness=0.9 mm with a $256 \times 256 \times 256$ acquisition matrix.

In addition, we used one of the 40 images from the LONI Probabilistic Brain Atlas (LPBA40) [26]. Each image had 56 different structures manually labeled, including a mask of the brain.

Table 2 shows the distance, overlap, and error metrics for the automated skull-stripping algorithms compared to manual segmentation. We compare BET,

BSE, and the Watershed method. Average metrics over the 630 subject images are shown. Our paired-sample t-tests showed the metric values across all the methods were significantly different. These test results showed that the deformable organisms approach was statistically better than the others in the Hausdorff distance and false negative error for our dataset.

The deformable organisms method took a few minutes to run on the subject images we used on a machine with dual 64-bit 2.4 gigahertz AMD Opteron 250 CPU with 8 gigabytes of memory.

We list average results of deformable organisms with and without error correction versus manual training in Table 3. Random samples of 20 images from the 630 were selected, using 10 to train and 10 to test the error correction. We repeat this 10 times and average the results. These average results were used a for two-sample t-test that found statistically significant improvement in Jaccard, Dice, and false positive error metrics.

Table 2: Distance, overlap, and error metrics comparing automated results with manual skull-stripping.

	Hausdorff Distance	Jaccard Overlap	Dice Overlap	False Negative Error	False Positive Error
Deformable Organisms	36.3475±24.3842	0.8478±0.0242	0.9175±0.0143	0.0253±0.0115	0.1328±0.0280
BET	41.7997±24.7972	0.8860±0.0183	0.9395±0.0104	0.0711±0.0218	0.0491±0.0189
BSE	64.7451±25.7714	0.8348±0.1295	0.9040±0.0842	0.1045±0.1456	0.0720±0.0323
Watershed	67.4672±8.7859	0.3650±0.0599	0.5321±0.0617	0.4511±0.0613	0.4831±0.0648

Table 3: Distance, overlap, and error metrics comparing the deformable organisms segmentation with and without error correction versus manual skull-stripping.

	Hausdorff Distance	Jaccard Overlap	Dice Overlap	False Negative Error	False Positive Error
Basic DO	35.4780±2.7890	0.8485±0.0009	0.9178±0.0005	0.0256±0.0006	0.1318±0.0010
DO with correction	35.5194±3.1299	0.8858±0.0013	0.9393±0.0007	0.0253±0.0004	0.0929±0.0012

7 Discussion

The metrics in Table 2 suggest that the performance of our deformable organisms approach is comparable to that of other widely used methods. It has the lowest Hausdorff distance average between its automatic results and the gold standard delineations. It did have a higher false positive error, meaning it may include slightly more voxels inside the boundary of the brain.

Table 3 shows that learning errors improves the segmentation results; all metrics examined were improved, especially the Jaccard coefficient. This additional training step may be useful if a large data set needs to be segmented, making it reasonable to segment some images manually for error correction. The method could be trained on a small subset of the manual and automatically segmented data, in a first pass, and the error corrections learned could be useful to segment the rest of the dataset.

Deformable organisms provide an adaptable framework to perform segmentation. They can encode a high-level plan into deformable models, to help them work together to accomplish segmentation tasks. The different control layers may be adapted to fit any type of segmentation tasks.

References

1. Boesen, K., Rehm, K., Schaper, K., Stoltzner, S., Woods, R., Luders, E., Rottenberg, D.: Quantitative comparison of four brain extraction algorithms. NeuroImage 22(3), 1255–1261 (2004)
2. Desbrun, M., Meyer, M., Schroder, P., Barr, A.: Implicit fairing of irregular meshes using diffusion and curvature flow. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. pp. 317–324. ACM Press/Addison-Wesley Publishing Co. (1999)
3. Freund, Y., Schapire, R.: A desicion-theoretic generalization of on-line learning and an application to boosting. In: Computational Learning Theory. pp. 23–37. Springer (1995)
4. Gering, D., Nabavi, A., Kikinis, R., Hata, N., ODonnell, L., Grimson, W., Jolesz, F., Black, P., Wells III, W.: An integrated visualization system for surgical planning and guidance using image fusion and an open MR. Journal of Magnetic Resonance Imaging 13, 967–975 (2001)
5. Gharachorloo, N., Gupta, S., Sproull, R., Sutherland, I.: A characterization of ten rasterization techniques. ACM SIGGRAPH Computer Graphics 23(3), 355–368 (1989)
6. Hamarneh, G., McIntosh, C.: Physics-based deformable organisms for medical image analysis. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. vol. 5747, pp. 326–335 (2005)
7. Huttenlocher, D., Klanderman, G., Rucklidge, W.: Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 850–863 (1993)
8. Ibanez, L., Schroeder, W., Ng, L., Cates, J., et al.: The ITK software guide. Kitware Inc (2005)
9. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Medical Image Analysis 5(2), 143–156 (2001)
10. Joshi, A., Lepore, N., Joshi, S., Lee, A., Barysheva, M., Stein, J., McMahon, K., Johnson, K., de Zubicaray, G., Martin, N., et al.: The contribution of genes to cortical thickness and volume. NeuroReport 22(3), 101 (2011)
11. Klein, A., Andersson, J., Ardekani, B., Ashburner, J., Avants, B., Chiang, M., Christensen, G., Collins, D., Gee, J., Hellier, P., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46(3), 786–802 (2009)
12. Lorensen, W., Cline, H.: Marching cubes: A high resolution 3D surface construction algorithm. ACM Siggraph Computer Graphics 21(4), 163–169 (1987)
13. McInerney, T., Hamarneh, G., Shenton, M., Terzopoulos, D.: Deformable organisms for automatic medical image analysis. Medical Image Analysis 6(3), 251–266 (2002)
14. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: a survey. Medical Image Analysis 1(2), 91–108 (1996)

15. McIntosh, C., Hamarneh, G.: I-DO: A Deformable Organisms framework for ITK. Medical Image Analysis (2006)
16. McIntosh, C., Hamarneh, G.: Spinal crawlers: Deformable organisms for spinal cord segmentation and analysis. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006 pp. 808–815 (2006)
17. McIntosh, C., Hamarneh, G.: Vessel crawlers: 3D physically-based deformable organisms for vasculature segmentation and analysis. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. vol. 1, pp. 1084–1091 (2006)
18. McIntosh, C., Hamarneh, G., Mori, G.: Human limb delineation and joint position recovery using localized boundary models. In: Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on. pp. 31–31. IEEE (2007)
19. Ohtake, Y., Belyaev, A., Bogaevski, I.: Polyhedral surface smoothing with simultaneous mesh regularization. In: Geometric Modeling and Processing 2000. Theory and Applications. Proceedings. pp. 229–237. IEEE (2000)
20. Pineda, J.: A parallel algorithm for polygon rasterization. ACM SIGGRAPH Computer Graphics 22(4), 17–20 (1988)
21. Prasad, G., Joshi, A., Thompson, P., Toga, A., Terzopoulos, D., Shattuck, D.: Skull-stripping with deformable organisms. In: International Symposium on Biomedical Imaging, ISBI 2011. pp. 1662–1665. IEEE (2011)
22. Resnick, S., Pham, D., Kraut, M., Zonderman, A., Davatzikos, C.: Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. Journal of Neuroscience 23(8), 3295–3301 (2003)
23. Roche, A., Malandain, G., Pennec, X., Ayache, N.: The correlation ratio as a new similarity measure for multimodal image registration. Medical Image Computing and Computer-Assisted Interventation–MICCAI 1998 p. 1115 (1998)
24. Segonne, F., Dale, A., Busa, E., Glessner, M., Salat, D., Hahn, H., Fischl, B.: A hybrid approach to the skull stripping problem in MRI. NeuroImage 22(3), 1060–1075 (2004)
25. Shattuck, D., Leahy, R.: BrainSuite: an automated cortical surface identification tool. Medical Image Analysis 6(2), 129–142 (2002)
26. Shattuck, D., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K., Poldrack, R., Bilder, R., Toga, A.: Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39(3), 1064–1080 (2008)
27. Shattuck, D., Prasad, G., Mirza, M., Narr, K., Toga, A.: Online resource for validation of brain segmentation methods. NeuroImage 45(2), 431–439 (2009)
28. Smith, S.: Fast robust automated brain extraction. Human Brain Mapping 17(3), 143–155 (2002)
29. Steels, L.: The artificial life roots of artificial intelligence. Artificial Life 1(1_2), 75–110 (1993)
30. Terzopoulos, D., Platt, J., Barr, A., Fleischer, K.: Elastically deformable models. ACM Siggraph Computer Graphics 21(4), 205–214 (1987)
31. Thompson, P., Hayashi, K., De Zubicaray, G., Janke, A., Rose, S., Semple, J., Herman, D., Hong, M., Dittmer, S., Doddrell, D., et al.: Dynamics of gray matter loss in Alzheimer's disease. Journal of Neuroscience 23(3), 994–1005 (2003)
32. Thompson, P., Hayashi, K., Sowell, E., Gogtay, N., Giedd, J., Rapoport, J., De Zubicaray, G., Janke, A., Rose, S., Semple, J., et al.: Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. NeuroImage 23, S2–S18 (2004)

33. Wang, H., Das, S., Suh, J., Altinay, M., Pluta, J., Craige, C., Avants, B., Yushkevich, P.: A Learning-Based Wrapper Method to Correct Systematic Errors in Automatic Image Segmentation: Consistently Improved Performance in Hippocampus, Cortex and Brain Segmentation. NeuroImage (2011)
34. Woods, R., Dapretto, M., Sicotte, N., Toga, A., Mazziotta, J.: Creation and use of a Talairach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data. Human Brain Mapping 8(2-3), 73–79 (1999)

Selecting Scales by Multiple Kernel Learning for Shape Diffusion Analysis

Umberto Castellani[1⋆], Aydın Ulaş[1], and Vittorio Murino[1,2], Marcella Bellani[3], Gianluca Rambaldelli[3], Michele Tansella[3], Paolo Brambilla[3,4]

[1] Department of Computer Science, University of Verona, Italy
[2] Istituto Italiano di Tecnologia, Genova, Italy
[3] Department of Medicine and Public Health, University of Verona, Italy
[4] ICBN Center, University of Udine and Verona, Italy

Abstract. Brain morphological abnormalities can typically be detected by advanced geometrical shape analysis techniques. Recently, shape diffusion methods have proved to be very effective in providing useful descriptions for brain classification purposes. In particular, they allow the analysis of such shapes at multiple scales, but the selection of the correct range of scales remains an open issue heavily affecting the performance of methods, and it needs to be estimated adaptively for different classes of shapes. In this paper, we focus on the diffusion scale selection in order to define a robust shape descriptor for brain classification. To this end, geometric features are extracted for each scale and the best feature combination is selected by employing *multiple kernel learning* (MKL). In the presented experiments, we compare the shape of Thalamic regions in order to discriminate between normal subjects and schizophrenic patients. We demonstrate that MKL allows to obtain classifiers which are more accurate with respect to other competing algorithms for schizophrenia detection. Moreover, using the weights computed by the MKL algorithm, we can select at which scale the features are more effective for schizophrenia classification.

Keywords: multiple kernel learning, schizophrenia, heat kernel, spectral shape analysis, support vector machines

1 Introduction

Recent advances in geometric shape analysis have led to a larger diffusion of computational anatomy methods, aimed at characterizing or modeling the morphological variations of biological shapes. One of the typical applications is analyzing the anatomy of organs, known as being possibly affected by abnormalities due to a certain disease, of several persons in order to discriminate between normal and pathological subjects [11,1]. To this aim, effective shape analysis

⋆ Corresponding author.

techniques are crucial to extract geometric features with high discriminant properties. A wide class of methods are based on the encoding of the deformation which aligns a pair of subjects, but such an approach requires the solution of a complex problem due to the non-linear registration between different shapes. More recently, new methods have been proposed to encode the shape geometric properties into a *descriptor* which compactly represents the whole shape, and performing the comparison by computing similarities in the descriptor space without any registration procedure. Among the shape analysis methods, diffusion geometry approaches are very promising since they are able to capture *intrinsic* characteristic of the shape. More specifically local geometric properties are encoded by the so-called *Heat Kernel* [16] which exploits heat diffusion characteristics at a given scale. The general idea consists of gaining information about the neighborhood of a point on the shape by recording the dissipation of heat over time from that point onto the rest of the shape. The fixed time is very important since it allows to capture different kinds of information: *local* shape characteristics are highlighted through the behavior of heat diffusion over short time periods, and, conversely, *global* shape properties are observed while considering longer periods [16,10]. So doing, simply varying a single parameter (the time), it is possible to characterize the properties of a shape at different scales. In particular, the so called *Heat Kernel Signature*(HKS) [16] has been proposed to encode simultaneously the contribution of local features for a fixed set of scales into a single shape descriptor. This general approach has been succesfully applied for object retrieval [4] and brain classification [6]. However, the choice of the range of the time periods to be evaluated (i.e., the *scales*) is critical and depends on the considered shape. In fact, for a particular shape, some scales may be highly discriminative, while some other scales should encode useless information. In this paper, we propose a new approach for integrating and selecting the contribution of geometric features collected at different scales by utilzing a Multiple Kernel Learning (MKL) approach. In general, MKL algorithms can learn a weighted combination of different kernel functions able to exploit information coming from multiple sources. In our case, the different sources are the features extracted at different scales. Therefore, several kernels are computed (i.e., one kernel per scale) and a set of weights are estimated for the kernel combination. In this fashion, we can choose the most discriminative scales by selecting those associated to the highest weights, and viceversa. Moreover, kernel combination leads to a new similarity measure which increases the classification accuracy. It is important to note that in our appraoch we aim at selecting the best shape characteristics for classification purposes, hence, our selection is driven by the performance of a Support Vector Machine (SVM) classifier. We have applied our method for brain classification in schizophrenic subjects: we have adopted a Region of Interest (ROI)-based method by analysing the shape of the Thalamic region, employing a *volumetric*-heat kernel computed for each voxel of the MRI scan at different scales, as described in our previous work [6]. This paper improves [6] for both methodological aspects, by proposing the automatic scale selection procedure and promising results. The rest of the paper is organized as

follows. In Section 2, the basics on shape diffusion procedures are reported. Section 3 describes the Multiple Kernel Lerning strategy, and the proposed method is reported in Section 4. Results are shown in Section 5 and conclusions are finally drawn in Section 6.

2 Shape analysis by heat diffusion

Considering a shape M as a compact Riemannian manifold [5], the heat diffusion on shape[5] is defined by the *heat* equation:

$$(\Delta_M + \frac{\partial}{\partial t})u(t, \boldsymbol{m}) = 0; \tag{1}$$

where u is the distribution of heat on the surface, $\boldsymbol{m} \in M$, Δ_M is the *Laplace-Beltrami* operator which, for compact spaces, has discrete eigendecomposition of the form $\Delta_M = \lambda_i \phi_i$. In this way, the *heat kernel* has the following eigendecomposition:

$$h_t(\boldsymbol{m}, \boldsymbol{m}') = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(\boldsymbol{m}) \phi_i(\boldsymbol{m}'), \tag{2}$$

where λ_i and ϕ_i are the i^{th} eigenvalue and the i^{th} eigenfunction of the Laplace-Beltrami operator, respectively. The heat kernel $h_t(\boldsymbol{m}, \boldsymbol{m}')$ is the solution of the heat equation with initial point heat source in $\boldsymbol{m}$ at time $t = 0$, and heat value in ending point $\boldsymbol{m}' \in M$ after time t. The heat kernel is *isometric invariant*, it is *informative*, and *stable* [16].

In the case of volumetric representations, the volume is sampled by a regular Cartesian grid composed by voxels, which allows the use of standard Laplacian in R^3 as the Laplace-Beltrami operator. We use finite differences to evaluate the second derivative in each direction of the volume. The heat kernel on volumes is invariant to volume isometries, in which shortest paths between points inside the shape do not change. Note that in real applications exact volume isometries are limited to the set of rigid transformations [15], however, also non-rigid deformations can faithfully be modelled as approximated volume isometries in practice. It is also worth noting that, as observed in [16,15], for small t the autodiffusion heat kernel $h_t(\boldsymbol{m}, \boldsymbol{m})$ of a point $\boldsymbol{m}$ with itself is directly related to the *scalar* curvature $s(\boldsymbol{m})$ [15]. More formally:

$$h_t(\boldsymbol{m}, \boldsymbol{m}) = (4\pi t)^{-3/2}(1 + \frac{1}{6}s(\boldsymbol{m})). \tag{3}$$

In practice, Equation 3 states that the heat tends to diffuse slower at points with positive curvature, and viceversa. This gives an intuitive explanation about the geometric properties of $h_t(\boldsymbol{m}, \boldsymbol{m})$, and suggests the idea of using it to build a shape descriptor [16].

[5] In this section, we borrow the notation from [16,5]

3 Multiple Kernel Learning

The main idea behind kernel methods [17] is to transform the input feature space to another space (eventually with a larger dimension) where the classes are linearly separable. In particular, by employing the SVM classifier, the discriminant function after the training phase becomes $f(x) = \langle \boldsymbol{w}, \Phi(\boldsymbol{x}) \rangle + b$, where $\boldsymbol{w}$ and b are the parameters of the hyperplane which separates two classes, and $\Phi(\cdot)$ is the mapping function. Using the dual formulation and the kernel trick, one does not have to define this mapping function explicitly and the discriminant function can be written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{4}$$

where $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle$ is the kernel function that calculates a similarity metric between data instances.

More recently, MKL methods have been proposed [3,13] for learning a combination k_η of several kernels:

$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\eta}) = f_\eta(\{k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)_{m=1}^P\}; \boldsymbol{\eta}) \tag{5}$$

where the combination function f_η forms a single kernel from P base kernels using the parameters vector (i.e., weights) $\boldsymbol{\eta}$. Such new kernel must be a valid kernel[6] [9] and can be plugged in Equation 4 for classification purposes. Different kernel functions correspond to different notions of similarity and instead of searching which works best, the MKL method does the picking for us, or may use a combination of kernels. MKL also allows us to combine different representations, possibly coming from different sources or modalities.

There is significant work on the theory and application of MKL, and most of the proposed algorithms differ among them by the optimization method employed to estimate the weights and by the used combination rule [3,13,14]. In this paper we focus on *linear*-MKL methods [3,13], whose general formulation is defined as:

$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\eta}) = \sum_{m=1}^{P} \eta_m k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m) \tag{6}$$

with $\eta_m \in \mathbb{R}$. As a simplest combination approach, the so called *fixed rules* [9] use the combination function in Eq. (6) with all weights equally set to $\boldsymbol{\eta}_m = 1$. Similarly, the *mean*-rule takes the mean of the kernels by setting all $\boldsymbol{\eta}_m = 1/P$. Indeed, in the most general case the weights $\boldsymbol{\eta}_m$ are automatically estimated by a *learning by example* approach. More specifically, MKL methods search for a combination of kernels that maximizes a generalized performance measure (i.e, *maximum margin* classification errors [9]). To this aim, in the training phase, both MKL weights and SVM parameters are simultaneously estimated within the same optimization problem.

[6] The validity of the kernel depends by the combination function.

4 The Proposed Method

The proposed method can be summarized in the following main steps:

1. MRI data collection.
2. Feature extraction at multiple scales.
3. Learning weights and classifier by MKL.
4. Scale selection and performance evaluation.

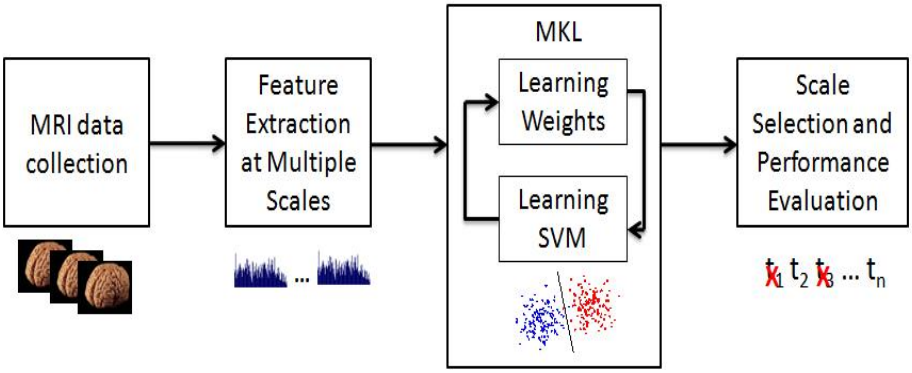


Fig. 1. General scheme of the proposed method.

MRI data collection. In order to employ a *learning-by-example* approach, we need a collection of samples for both healthy subjects and patients. Source data are MRI scans where shape information can be provided in terms of volumetric data.

Feature extraction at multiple scales. According to the shape diffusion analysis described in Section 2, for each subject geometric features are extracted at multiple scales: a set of time values $(t_1, t_2, \cdots, t_n)$ are defined, and the autodiffusion value is computed for each voxel $\boldsymbol{m}$, leading to:

$$H_{t_i}(M) = \{h_{t_i}(\boldsymbol{m}, \boldsymbol{m}), \forall \boldsymbol{m} \in M\}.$$

Then, such values are accumulated into a histogram $r_i = hist(H_{t_i}(M))$. In this manner, we obtain a set of n sources of shape representation $\{r_1, \cdots, r_n\}$, each one encoding the global shape at a certain scale. The number of bins for each histogram is chosen as 100.

Learning weights and classifier by MKL. The contribution of geometric features extracted at each scale are combined by employing the MKL strategy as described in 3. Each shape representation r_i is associated to a kernel k_m by leading to $n = P$ kernels. Indeed, both the weights $(\eta_1 \cdots \eta_P)$ and the SVM parameters are estimated. In order to obtain the best classification accuracy according to the *max-margin* paradigm an *alternating* approach is used between the optimization of kernel weights and the optimization of the SVM classifier. In each step, given the current solution of kernel weights, MKL solves a standard SVM optimization problem with the combined kernel. Then, a specific procedure is applied to update the kernel weights.

Scale selection and performance evaluation. Once the MKL procedure is completed, we obtain a two-fold advantage: i) we can select the best scale contributions by keeping only the scales associated to the highest weighs, and ii) we can compose a new kernel from the weighted contributions of the best scales, which can be evaluated for classification purposes.

5 Experiments

This section is organized in the following parts: i) data gathering, ii) experimental methodology, iii) results, and iv) discussion.

5.1 Data Gathering

Quantitative data collection and processing in MRI-based research implies to face several methodological issues to minimize biases and distortions. The standard approach to deal with these issues is following well-established guidelines dictated by international organizations, such as the World Health Organization (WHO), or codified by respected institutions, such as leading universities. All patients received a diagnosis of schizophrenia according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders [2]. In this work, we employ a ROI-based approach [11], so only a well defined brain subpart has been considered in this study. More specifically, we focus our analysis on the left-Thalamus whose abnormal activity has been already investigated in schizophrenia[8]. ROIs have been manually traced by experts, according to well defined medical protocols. The data set used in this work is composed by MRI brain scans of 30 patients affected by schizophrenia and 30 healthy control subjects.

5.2 Experimental protocol

In our experiments, we apply leave-one-out (LOO) cross-validation to assess the performance of the technique. Since LOO is used as the cross validation technique, we do not report standard deviations or variances. We compare our results using k-fold paired t-test at $p = 0.05$. We collect geometric features at 11 scales generating different shape representations $r_{01}, \cdots, r_{11}$. In practice, each

representation r_i is a feature vector x_i which is plugged in the MKL framework. We employ the dot product as basic kernel function (i.e., linear kernel) since it avoids the estimation of free kernel parameters. Different strategies to combine the different shape representations have also been evaluated:

- Single Best Kernel (`Single-best`): an SVM is trained separately per each representation. Therefore, the performance of the classification are evaluated separately at each scale. So doing, we can evaluate the independent contributions coming from the different sources of information and select the best one.
- Feature concatenation (`SVM-con`): the contributions coming from the different sources are concatenated into a single feature vector. Then, a single SVM is employed for classification [7].
- Rule-based MKL (`RBMKL`): as baseline MKL approach, the so called rule-based method is evaluated: the kernels computed at each scale are combined by simply taking their average (i.e., $\forall m, \boldsymbol{\eta_m} = 1/P$).
- Simple MKL (`SimpleMKL`): a simple but effective MKL algorithm is employed [14] by addressing the MKL problem through a weighted 2-norm regularization formulation with additional constraint on the weights that encourages sparse kernel combination. It is a popular approach and its code is publicly available[8].
- Group Lasso MKL (`GLMKL`): it denotes the group Lasso-based MKL algorithms proposed by [12,18]. A closed form solution for optimazing the kernel weights based on the equivalence between group-lasso and MKL is proposed. In our implementation, we used l_1-norm on the kernel weights and learned a convex combination of the kernels.

5.3 Results

The first evaluation scores are shown in Table 1, which reports the single-best kernel accuracies for all feature representations. We can observe that the best performance is obtained at 78.33 % using `r02` which is shown as bold face in the table. The entries marked with "*" show the accuracies which are statistically significantly less accurate than the best algorithm using k-fold paired t-test at $p = 0.05$.

Table 1. Single-kernel SVM accuracies.

r01	r02	r03	r04	r05	r06	r07	r08	r09	r10	r11
75.00	**78.33**	76.67	76.67	73.33	*66.67	68.33	70.00	76.67	71.67	70.00

Second, concatenating the features in a single vector leads to 83.33 % accuracy.

[7] We use LIBSVM software [7] to train the SVM.
[8] http://asi.insa-rouen.fr

Third, using the proposed three different MKL algorithms, we combined the eleven kernels by introducing the weights η_m. Table 2 reports the results of the best single-kernel SVM, the accuracy of the concatenated feature set, and the three MKL-based algorithms trained. The values in parantheses show the percentage of controls classified as schizophrenia and the percentage of patients classified as healthy respectively. We achieve an accuracy of 86.67%, reached by combining eleven kernels with the `SimpleMKL` approach. This result is better than all other MKL settings and single-kernel SVMs. Further, `GLMKL` achieves 85% accuracy which is still higher than that reached by the feature concatenation method. We can also note that we cannot overcome `SVM-con` when we use `RBMKL`, as the latter gives equal weight to each kernel. In fact, if there are inaccurate representations in the given set, the overall mean combination accuracy may be less of that reached using the single best. Conversely, when the weights are automatically estimated, such as in `SimpleMKL` and `GLMKL` the selection of the most reliable information is carried out by the MKL procedure and the overal performance improves.

Table 2. MKL accuracies (false positives and negatives are reported in brackets).

`Single-best`	`SVM-con`	`RBMKL`	`SimpleMKL`	`GLMKL`
*78.33 (10, 11.6)	83.33 (8.3, 8.3)	*81.67 (10, 8.3)	**86.67** (6.6, 6.6)	85.00 (8.3, 6.6)

In Figure 2, we plotted the weights of MKL for both `SimpleMKL` and `GLMKL` algorithms. Note that the estimated weights are coherent in the two algorithms. As expected, the best representation is `r02`, which has the highest weights. Although the other representations with high weights (`r08`, `r11` and `r05`) do not provide much accurate single-kernel SVMs results, their contributions to the overall accuracy in the combination is higher than those given by the other kernels. This demonstrates that when considering combinations, even a representation which does not lead to very precise results may contribute to raise the overall combination accuracy. Moreover, we can also deduce that these four representations are the most useful in discriminating between healthy and schizophrenic subjects, and we may focus the attention on these properties only.

Using this information, we also performed the above pipeline using only these four representations, and we can observe the results in Table 3. Using this subset, we get the highest accuracy with `SimpleMKL`[9], reaching 88.33% of accuracy. We can also observe an increase in `RBMKL`.

[9] Note that in principle the same result should have been obtained automatically from MKL algorithms on all representations. In practice, this is not the case in our experiment due to the fact that the estimated solution is trapped into a local minimum.

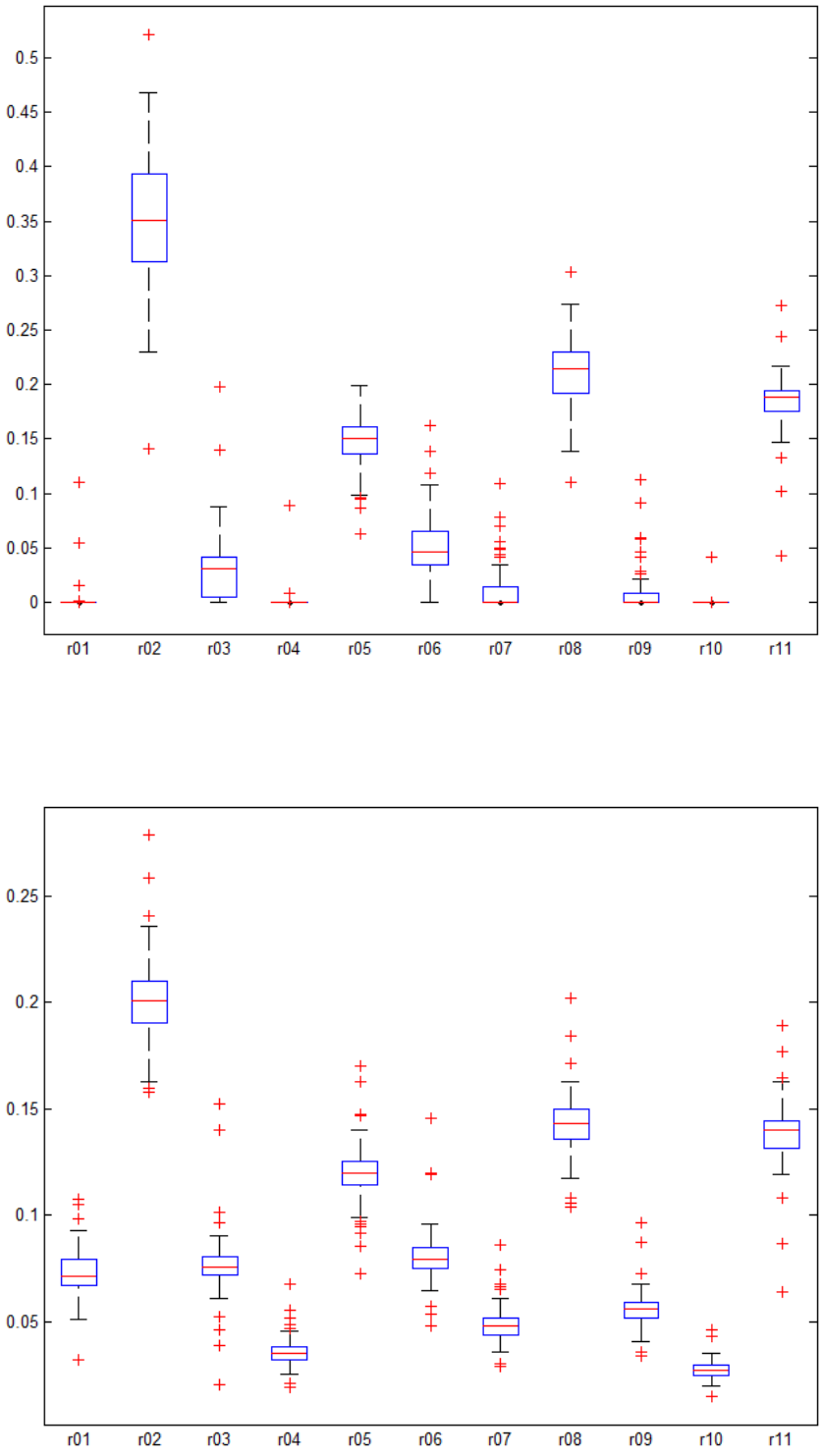


Fig. 2. Combination weights in MKL using the linear kernel. Top: using `SimpleMKL`, Bottom: using `GLMKL`.

Table 3. MKL accuracies on the selected subset of representations (false positives and negatives are reported in brackets).

SVM	SVM-con	RBMKL	SimpleMKL	GLMKL
*78.33 (10, 11.6)	*83.33 (6.6, 10)	*83.33 (6.6, 10)	**88.33** (6.6, 5)	85.00 (6.6, 8.3)

5.4 Discussion

In this work, we have shown in general that MKL algorithms perform better than both single-best kernel SVMs and feature concatenation strategies. We have also observed that `RBMKL` (which does not compute weights while combining kernels) does not outperform the feature concatenatenation approach. Conversely, when the kernel combination is carried out by estimating proper weights, a drastic improvement is instead obtained. The kernel weights also allow us to extract useful information: it is interesting to observe that, for both MKL algorithms with the highest accuracy, four representations have the maximum effect (i.e., the highest weights), i.e., `r02`, `r08`, `r11`, and `r05`, with `r02` being the best single-kernel. We use this information to select a smaller number of representations to reduce the costs of the feature extraction phase. Finally, we can also observe that by using such subset we can reach the best accuracy overall.

6 Conclusions

In this paper, we focus on scale selection for anatomical shape characterization. By employing a shape diffusion approach, we extract several shape descriptors at different scales in order to discriminate between healthy subjects and patients affected by schizophrenia. We have shown that machine learning techniques can be useful to improve the shape analysis in these (biomedical) contexts. We propose a Multiple Kernel Learning algorithm for the automatic estimation of the best feature representation for classification purposes. In this way, being driven by the training data, we are able to choose the scales of the heat kernel which are more suitable to describe our kind of shapes. In particular, in our experiments addressing the Thalamic region classification, we have shown that both small and high scales are crucial. Actually, the best accuracy is observed at `r02` for which very local information are collected from the shape. Nevertheless, when also higher scales are considered the performance is further improved, meaning that also global shape information is relevant.

Acknowledgements

We acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (contract 213250). We thank Dr. Mehmet Gönen for the implementation of the MKL algorithms.

References

1. Agarwal, N., Port, J.D., Bazzocchi, M., Renshaw, P.F.: Update on the use of MR for assessment and diagnosis of psychiatric diseases. Radiology 255(1), 23–41 (2010)
2. American Psychiatric Association: Diagnostic and statistical manual of mental disorders, DSM-IV. Washington DC, 4th edn. (1994)
3. Bach, F.R., Lanckriet, G.R.G., Jordan, M.I.: Multiple kernel learning, conic duality, and the smo algorithm. In: Proceedings of the twenty-first international conference on Machine learning, ICML '04. pp. 41–48 (2004)
4. Bronstein, A.M., Bronstein, M.M., Ovsjanikov, M., Guibas, L.J.: Shape Google: geometric words and expressions for invariant shape retrieval. ACM Transaction on Graphics 30(1), 1–20 (2011)
5. Bronstein, A.M., Bronstein, M.M., Ovsjanikov, M., Guibas, L.J.: Shape recognition with spectral distances. IEEE Trans. Pattern Analysis and Machine Intelligence 33(5), 1065–1071 (2011)
6. Castellani, U., Mirtuono, P., Murino, V., Bellani, M., Rambaldelli, G., Tansella, M., Brambilla, P.: A new shape diffusion descriptor for brain classification. In: Medical Image Comp. Computer-Assisted Intervention (MICCAI) (2011)
7. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), `http://www.csie.ntu.edu.tw/$\sim$cjlin/libsvm`
8. Corradi-DellAcqua, C., Tomelleri, L., Bellani, M., Rambaldelli, G., Cerini, R., Pozzi-Mucelli, R., Balestrieri, M., Tansella, M., Brambilla, P.: Thalamic-insular dysconnectivity in schizophrenia: Evidence from structural equation modeling. Human Brain Mapping p. in press (2011)
9. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and other Kernel-based Learning Methods. Cambridge University Press (2000)
10. Gebal, K., Baerentzen, J.A., Aanaes, H., Larsen, R.: Shape analysis using the auto diffusion function. In: In SGP (2009)
11. Giuliani, N.R., Calhouna, V.D., Pearlson, G.D., Francis, A., Buchanan, R.W.: Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. Schizophrenia Research 74(2–3), 135–147 (2005)
12. Kloft, M., Brefeld, U., Sonnenburg, S., Zien, A.: l_p-norm multiple kernel learning. Journal of Machine Learning Research 12, 953–997 (2011)
13. Lanckriet, G.R.G., Cristianini, N., Bartlett, P., Ghaoui, L.E., Jordan, M.I.: Learning the kernel matrix with semidefinite programming. Journal of Machine Learning Research 5, 27–72 (December 2004)
14. Rakotomamonjy, A., Bach, F., Canu, S., Grandvalet, Y.: SimpleMKL. Journal of Machine Learning Research 9, 2491–2521 (2008)
15. Raviv, D., Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Volumetric heat kernel signatures. In: Workshop on 3D Object Retrieval (2010)
16. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: Proceedings of the Symposium on Geometry Processing. pp. 1383–1392 (2009)
17. Vapnik, V.N.: Statistical learning theory. John Wiley and Sons (1998)
18. Xu, Z., Jin, R., Yang, H., King, I., Lyu, M.R.: Simple and efficient multiple kernel learning by group Lasso. In: Proceedings of the 27th International Conference on Machine Learning, ICML '10. pp. 1175–1182 (2010)

Probabilistic Modeling of Landmark Distances and Structure for Anomaly-proof Landmark Detection

Shouhei Hanaoka[1], Yoshitaka Masutani[1,2], Mitsutaka Nemoto[1], Yukihiro Nomura[1], Takeharu Yoshikawa[3], Naoto Hayashi[3], Naoki Yoshioka[4], Kuni Ohtomo[1,2]

[1] Department of Radiology, [2] Division of Radiology and Biomedical Engineering, Graduate School of Medicine, [3] Department of Computational Diagnostic Radiology and Preventive Medicine and [4] Department of Integrated Imaging Informatics, The University of Tokyo hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
hanaoka-tky@umin.ac.jp

Abstract. A combinatorial optimization algorithm for detecting multiple anatomical landmarks is presented. It can determine the positions of over 100 landmarks concurrently, taking spatial correlations of all landmark pairs into account. Provided that a set of landmark candidate lists is given by sensitivity-optimized single-landmark detectors, the proposed algorithm can find the most probable combination of them through solving a MAP estimation-based combinatorial optimization problem. Additionally, it is designed to handle subjects with "segmentation anomaly of the spinal column," a common anatomical anomaly of the spine. The proposed system was evaluated with 156 landmarks in 50 datasets, using virtually created detector output sets. In the result, the algorithm achieved 97.6% of spinal anomaly estimation accuracy even with 50 points of candidates given per landmark, as well as 96.2% of accuracy in landmark candidate selection. From these results, usefulness of the proposed algorithm for subjects with spinal anomaly was suggested.

Keywords: Landmark, Combinatorial optimization, MAP estimation, Anatomical anomaly, Computed tomography, Spine

1 Introduction

Landmark point detection algorithms are extensively researched and widely used in various medical image processing applications. However, it is a difficult task to detect a large number of landmarks correctly, because the human body includes a lot of similar regions sharing their appearances. Even most of frequently-used landmarks do not have truly unique local shape or intensity. Consequently, detection results often include a certain number of false positive candidates.

Furthermore, some sort of important anatomical landmarks can intrinsically be non-existent, not only in patients due to pathological condition, but even in healthy subjects due to anatomical anomalies (fig. 1a). Confirming such a situation is very difficult by detecting each landmark independently and sequentially.

One solution is to determine the entire landmark positions simultaneously, taking their spatial relation into account. It can be done by dividing the whole problem into 2 sequential phases: the individual landmark detection phase and the combinatorial optimization phase (figs. 1b, 1c). In the former phase, each landmark is detected by a single detector. Each detector is optimized to maximize the sensitivity, not the specificity, so the detection result forms a candidate list which includes a lot of false positives. In the latter phase, the best combination of choice from all candidate lists is selected with the use of *a priori* knowledge on the inter-landmark relationship. Defect of any landmark can also be detected through this optimization phase.

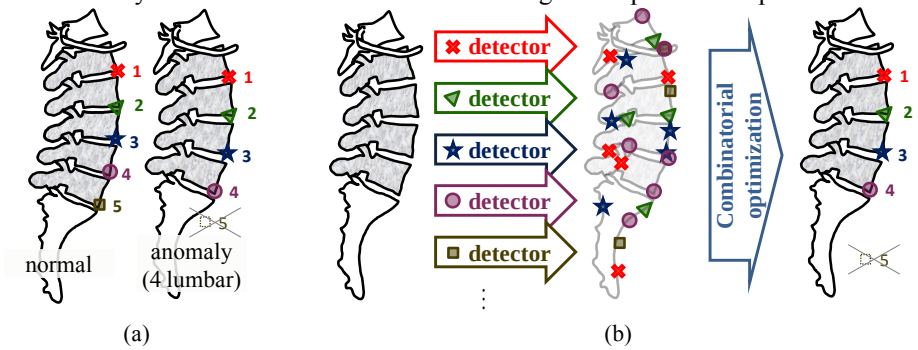


Fig. 1. (a) An example of anatomical anomaly and landmark deficit in the 5th lumbar vertebra. (b) A schema of the framework of the proposed method. Note that, for the non-existent landmark (□), all detected candidates are rejected by the following combinatorial optimization.

Among related studies, Seifert et al. reported a framework to automatically detect 19 landmarks. [1] In their approach, a belief propagation algorithm is used with prior knowledge about landmarks' geometric relationships such as "to the right of", "close to", etc. Though their result was excellent, the prior knowledge used and the way to build it were not described precisely.

In this paper, we introduce a framework to determine over 100 landmark positions reliably. Especially, we focus upon our method to solve the landmark-set combinatorial optimization problem with use of maximum *a posteriori* (MAP) estimation. The method was evaluated with 156 landmarks in 50 human body CT image datasets. Virtually-created detector outputs were used in this study, rather than those of real detectors [2], in order to evaluate the ability of our framework to handle a large number of false candidates. Furthermore, handling of common anomaly (alteration in the number of thoracic/lumbar vertebrae) was also evaluated.

2 Methods

2.1. Definition

2.1.1. Detector output. Firstly, we defined a landmark detector generally. In our framework, each detector has to output not only a series of candidate positions, but also estimated probabilities of them. Based on this, outputs from one detector (for the *m*-th landmark) are defined as:

N_m	number of the candidates	$(0 \le N_m)$
$\mathbf{c}_m^{\,i}$	coordinates of the *i*-th candidate	$(1 \le i \le N_m)$
$p_m^{\,i}$	detector-estimated probability of the *i*-th candidate	$(0 < p_m^{\,i} \le 1)$
$p_m^{\,\Im}$	probability that "none of candidates $\mathbf{c}_m^{\,i}$ is correct"	$(0 \le p_m^{\,\Im} \le 1)$

The whole set of output is also defined as $\mathrm{I}_m = \left\{N_m, \mathbf{c}_m^{\,1}, \ldots, \mathbf{c}_m^{\,N_m}, p_m^{\,1}, \ldots, p_m^{\,N_m}, p_m^{\,\Im}\right\}$. Any output I_m must satisfy $p_m^{\,\Im} + \sum p_m^{\,i} = 1$.

Letting the true landmark position be $\mathbf{x}_m$, the detector-estimated probability $p_m^{\,i}$ can be interpreted as a conditional probability with a certain detector output I_m.

$$p_m^{\,i} = p\left(\mathbf{x}_m = \mathbf{c}_m^{\,i} \mid \mathrm{I}_m\right) \tag{1}$$

If $p_m^{\,\Im} > 0$, it indicates that there are some possibilities of "no true landmark position is included in the candidate set $\left\{\mathbf{c}_m^{\,1}, \ldots, \mathbf{c}_m^{\,N_m}\right\}$." The proposed algorithm can consider such a situation, and handles it as one extra state which will be represented as "$\mathbf{x}_m = \Im$" in this paper. The state $\Im$ can be a "true answer" when the target landmark is out of the imaging range, or corresponding landmark does not exist anywhere (e.g., due to some anatomical anomaly or pathological condition). In this describing method, the probability $p_m^{\,\Im}$ can also be interpreted as a conditional probability as

$$p_m^{\,\Im} = p\left(\mathbf{x}_m = \Im \mid \mathrm{I}_m\right). \tag{2}$$

2.1.2. The prior probability distribution of LM positions. Secondly, the prior probability function for all possible landmark position sets has to be defined in advance.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ be the positional vectors of total *M* landmarks, and $\mathbf{X} = (\mathbf{x}_1^{\,t}\, \mathbf{x}_2^{\,t} \ldots \mathbf{x}_M^{\,t})^t$ be the concatenated form of them. Each element $\mathbf{x}_m$ can be regarded as a stochastic variable whose domain is $\mathbf{x}_m \in \left\{\Im, \mathbf{c}_m^{\,1}, \ldots, \mathbf{c}_m^{\,N_m}\right\}$. The aim is to approximate the prior probability distribution of $\mathbf{X}$ as a single function $p(\mathbf{X})$, in order to use in the following MAP estimation.

In this study $p(\mathbf{X})$ is defined as a function of the squared distances between all landmark pairs. When any $\mathbf{x}_m$ $(1 \le m \le M)$ satisfies $\mathbf{x}_m = \Im$, however, the corresponding term of $p(\mathbf{X})$ is replaced by a conventionally-defined constant term. The calculation method of $p(\mathbf{X})$ without considering $\Im$ is discussed in the next chapter, which will be followed by the general definition of $p(\mathbf{X})$.

(i) *When* $\mathbf{x}_m \neq \Im, \forall m$. Let

$$d_{i,j} = \left|\mathbf{x}_i - \mathbf{x}_j\right|^2 \quad (1 \le i < j \le M) \tag{3}$$

be the squared distance between the i-th and j-th landmarks. Note that $d_{i,j}$ is defined only if both $\mathbf{x}_i$ and $\mathbf{x}_j$ are not $\Im$. Then, the distance is normalized by its average $\mathrm{E}(d_{i,j})$ and variance $\mathrm{V}(d_{i,j})$ in the training datasets. The normalized distance $g_{i,j}$ follows the equation

$$g_{i,j} = \frac{d_{i,j} - \mathrm{E}(d_{i,j})}{\sqrt{\mathrm{V}(d_{i,j})}} \qquad \text{if } \mathbf{x}_i \neq \Im, \mathbf{x}_j \neq \Im. \qquad (1 \le i < j \le M) \tag{4}$$

Let a vector $\mathbf{G} = (g_{1,2}\ g_{1,3} \cdots\ g_{i,j} \cdots g_{M-1,M})^t$ be the concatenated normalized squared distances between all landmark pairs. Note that the vector $\mathbf{G}$ has ${}_M\mathrm{C}_2 = \frac{M(M-1)}{2}$ of elements. Then, the prior probability distribution $p(\mathbf{X})$ was approximated by a multivariate normal distribution of the vector $\mathbf{G}$. That is,

$$p(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^{{}_M\mathrm{C}_2} \cdot \sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2}\mathbf{G}^t\mathbf{V}^{-1}\mathbf{G}\right) \tag{5}$$

where $\mathbf{V}$ represents the covariance matrix of $\mathbf{G}$, which is calculated from training sets, and $|\mathbf{V}|$ is the determinant of $\mathbf{V}$. The size of the matrix $\mathbf{V}$ is ${}_M\mathrm{C}_2 \times {}_M\mathrm{C}_2$.

(ii) *General definition*. Eq. (5) can be written in an extended form

$$p(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^{{}_M\mathrm{C}_2} \cdot \sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\sum_{k=1}^{M-1}\sum_{l=k+1}^{M} g_{i,j} \cdot \left\{\mathbf{V}^{-1}\right\}_{(i,j),(k,l)} \cdot g_{k,l}\right) \tag{6}$$

where $\{\mathbf{V}^{-1}\}_{(i,j),(k,l)}$ is the corresponding element of the matrix $\mathbf{V}^{-1}$. Here, the summed term $g_{i,j} \cdot \left\{\mathbf{V}^{-1}\right\}_{(i,j),(k,l)} \cdot g_{k,l}$ is not available when any of $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$ or $\mathbf{x}_l$ is $\Im$. In order to define $p(\mathbf{X})$ generally, we replaced them as follows:

$$p(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^{{}_M\mathrm{C}_2} \cdot \sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\sum_{k=1}^{M-1}\sum_{l=k+1}^{M} t_{(i,j),(k,l)}\right) \tag{7}$$

$$t_{(i,j),(k,l)} = \begin{cases} g_{i,j} \cdot \left\{\mathbf{V}^{-1}\right\}_{(i,j),(k,l)} \cdot g_{k,l} & \text{if all of } \mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k \text{ and } \mathbf{x}_l \text{ are not } \Im \\ 0 & \text{otherwise, and } (i,j) \neq (k,l) \\ \gamma & \text{otherwise, and } (i,j) = (k,l) \end{cases}$$

When $(i,j) \neq (k,l)$, the summed term $t_{(i,j),(k,l)}$ evaluates how $g_{i,j}$ and $g_{k,l}$ are correlated as like that of training sets. When no prior information on $g_{i,j}$ (or $g_{k,l}$) is available, the expected value of this term should be zero. That is why it should be replaced by zero in case of $\Im$.

On the other hand, when $(i,j)=(k,l)$, the term $t_{(i,j),(i,j)} = \{\mathbf{V}^{-1}\}_{(i,j),(i,j)} \cdot g_{i,j}{}^2$ always has a positive value which evaluates how the distance $d_{i,j}$ varies from that of the training sets. Therefore, we replaced it by a positive constant γ. Because increasing this

parameter will reduce the probability $p(\mathbf{X})$ for any $\mathbf{x}_m=\Im$, γ can be regarded as a parameter which controls how the algorithm avoids the state $\Im$, or how it prefers any detected candidate $\mathbf{c}_m^{\,i}$, for each landmark. We empirically selected γ=2 based on our preliminary experiments (the data is not shown).

2.1.3. *a posteriori* probability distribution. Once the detector outputs are given, the probability *a posteriori* can be calculated with Bayes' theorem. From the series of detector outputs I_1, I_2, …, I_M and the prior probability distribution $p(\mathbf{X})$, the posterior probability can be calculated by the Bayes' as:

$$p(\mathbf{X} \mid \mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M) = \frac{p(\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M \mid \mathbf{X}) \cdot p(\mathbf{X})}{p(\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M)} \tag{8}$$

The denominator is constant and independent of $\mathbf{X}$. Therefore, the maximum *a posteriori* estimation of landmark position set $\mathbf{X}$ is as follows:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} p(\mathbf{X} \mid \mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M) = \arg\max_{\mathbf{X}} p(\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M \mid \mathbf{X}) \cdot p(\mathbf{X}) \tag{9}$$

We assumed that all detector outputs are independent of each other, as well as independent on the positions of the other landmarks. It means that each detector output is only dependent on the corresponding landmark position. Then the term $p(\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M \mid \mathbf{X})$ in (9) can be divided into the product of single-landmark conditional probabilities $p(\mathrm{I}_m \mid \mathbf{x}_m)$. That is,

$$p(\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M \mid \mathbf{X}) = \prod_{m=1}^{M} p(\mathrm{I}_m \mid \mathbf{x}_m) = \prod_{m=1}^{M} \frac{p(\mathbf{x}_m \mid \mathrm{I}_m) \cdot p(\mathrm{I}_m)}{p(\mathbf{x}_m)} . \tag{10}$$

The term $p(\mathrm{I}_m)$ in this formula is independent of $\mathbf{X}$, so it can be ignored through the MAP estimation. The denominator $p(\mathbf{x}_m)$ is also can be ignored, because the term is the probability distribution of one single landmark position $\mathbf{x}_m$ *without* any prior information. Ignoring the term is equivalent to regarding it to be constant and homogeneous anywhere. The remaining term $p(\mathbf{x}_m \mid \mathrm{I}_m)$ is the detector-estimated probability as described in Eqs. (1) and (2). That is,

$$p(\mathbf{x}_m \mid \mathrm{I}_m) = \begin{cases} p_m^{\,i} & \text{if } \mathbf{x}_m = \mathbf{c}_m^{\,i}, \exists i \\ p_m^{\,\Im} & \text{if } \mathbf{x}_m = \Im \end{cases} . \tag{11}$$

In conclusion, the MAP estimation will be performed by the following formula:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} p(\mathbf{X} \mid \mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, ..., \mathrm{I}_M) = \arg\max_{\mathbf{X}} \left[\left\{ \prod_{m=1}^{M} p(\mathbf{x}_m \mid \mathrm{I}_m) \right\} \cdot p(\mathbf{X}) \right] \tag{12}$$

2.2. Implementation

2.2.1. Tikhonov's regularization. In practice, the dimension of **G** in Eq. (5) (or Eq. (7)), which equals ${}_M\mathrm{C}_2 = \frac{M(M-1)}{2}$, can be much greater than the number of training cases N. In such a case, the covariance matrix **V** will be rank-deficit and has no inverse matrix $\mathbf{V}^{-1}$. To avoid them, Tikhonov's regularization [3] was performed by replacing **V** with its regularized matrix $\mathbf{V}_{\mathrm{reg}}$ as follows:

$$\mathbf{V}_{reg} = \mathbf{V} + \lambda \mathbf{I} \tag{13}$$

We empirically selected $\lambda = 1.0$ for this study.

Note that the calculation of Eq. (5) can be speeded up by using the following formula (derived from the Woodbury matrix identity),

$$\mathbf{V}_{reg}^{-1} = \left(\mathbf{V} + \lambda \mathbf{I}\right)^{-1} = \left(\mathbf{U}^t \mathbf{D} \mathbf{U} + \lambda \mathbf{I}\right)^{-1} = \lambda^{-1}\mathbf{I} - \lambda^{-2} \cdot \mathbf{U}^t \left(\mathbf{D}^{-1} + \lambda^{-1} \cdot \mathbf{I}'\right)^{-1} \mathbf{U} \tag{14}$$

Here, **D** is a $(N\text{-}1)\times(N\text{-}1)$ diagonal matrix whose diagonal elements are nonzero eigenvalues of **V**, and **U** is a $(N\text{-}1)\times {}_M\mathrm{C}_2$ matrix whose row vectors are corresponding eigenvectors. **I** and **I'** are identity matrices whose sizes are ${}_M\mathrm{C}_2$ and N-1, respectively.

2.2.2. Combinatorial optimization. The maximization of Eq. (12) for all combinations of $\mathbf{X}=\{\mathbf{x}_m\}$, $\mathbf{x}_m \in \left\{\Im, \mathbf{c}_m^{\,1}, \ldots, \mathbf{c}_m^{\,N_m}\right\}$ is performed by a Gibbs' sampler-based simulated annealing algorithm reported by Geman & Geman [4]. In the algorithm, a virtual temperature T is introduced to modify the probability distribution. The distribution in (12) was modified as follows:

$$p_{\mathrm{modified}}(\mathbf{X};T) = \frac{1}{Z} \cdot \left[\left\{ \prod_{m=1}^{M} p(\mathbf{x}_m \mid \mathrm{I}_m) \right\} \cdot p(\mathbf{X}) \right]^{-\frac{1}{T}} \tag{15}$$

Z is a normalization factor in order to make the sum of probability 1.

When T=1, the modified distribution equals to the original in (12). With higher T, the modified distribution is almost homogenous for any **X** in its domain. However, it becomes sharper and more local with lower T. In the limiting case of $T \to +0$, the modified distribution has nonzero probability only at its global maximum point. (Fig. 2)

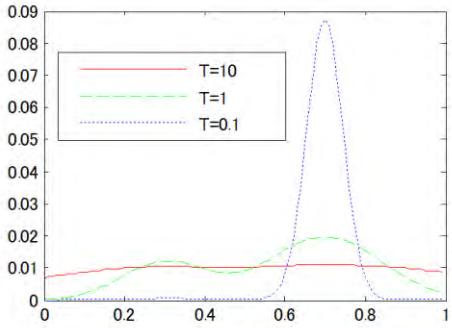


Fig. 2. A simple example how the probability distribution is modified by the temperature. The green curve illustrates an original distribution (a unary mixture Gaussian distribution for example). The modified distributions with T = 10 and 0.1 are illustrated with red and blue curves, respectively. Note that the distribution becomes flat with higher temperature, while it becomes sharper with lower temperature.

In the simulated annealing, each $\mathbf{x}_m$ (m = 1, 2, …, M) is sequentially and repeatedly sampled from the modified distribution by a Gibbs' sampler. The sampling begins with very high T, which decreases gradually, and finally it gets so low that the system converges to the maximum point. In this study the simulated annealing were performed for 100 cycles with T=1000 (in order to cancel the effect of the initial condition; so-called *burn-in*), then it was gradually cooled down to T=0.01 through 1000 cycles.

2.2.3. Handling of vertebral anomalies. A majority of human beings have 12 thoracic and 5 lumbar vertebrae. Segmentation anomaly of the spinal column is a common anatomical anomaly in which the subject has 11 or 13 thoracic, and/or 4 or 6 lumbar, vertebrae. The prevalence is, in a report, about 9 % [5]. This anomaly is very problematic in both defining and detecting vertebral landmarks (fig. 1a). The "state ℑ" approach described in chapter 2.1. is not enough for them, because it is not only a local banishment of a single anatomical entity but causing a global morphological change in the spine.

To overcome this, a series of "anomaly landmark position set converters" are introduced. One converter can convert any landmark position set in a subject with a certain type of anomaly (e.g., 6 lumbar vertebrae, or "6L") into a virtually normalized landmark position set (i.e., a landmark position set as if she or he has only 5 lumbar vertebrae). It is simply performed by replacing each landmark coordinates by an appropriate internally dividing point between two of them (fig. 3). Because 7 types of anomalies (11T, 13T, 4L, 6L, 11T+6L, 13T+4L) were considered in this study, 7 different converters were designed. Through one of these converters, abnormal spines can be converted into a "normalized" one, with which the prior probability $p(\mathbf{X})$ can be calculated.

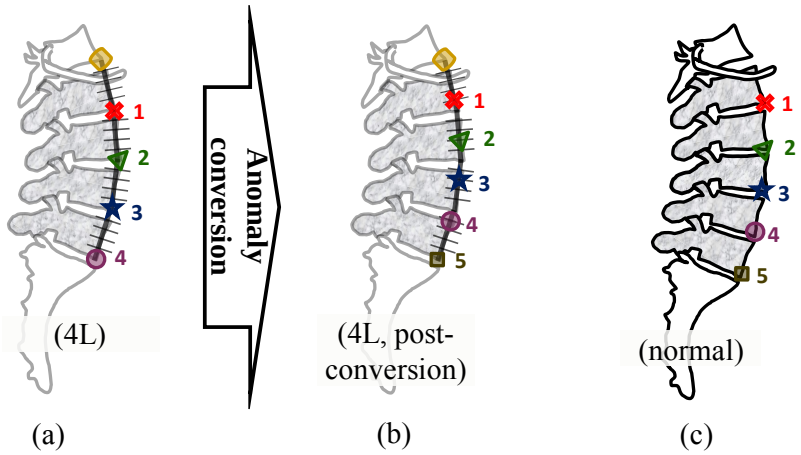


Fig. 3. The anomaly conversion. In this example, (a) a landmark position set in a case of 4-lumbar vertebra (4L) anomaly is virtually converted into (b) a "normal" landmark set with 5 vertebrae. With this conversion, the posterior probability $p(\mathbf{X})$ of the given landmark position set can be calculated in the same manner as in the cases with (c) normal 5-lumbar spines.

Because the algorithm do not know which anomaly is correct, it must estimate it. It can be done by (i) hypothesizing one anomaly (or normal), (ii) calculating the posterior probability under the hypothesis and (iii) comparing the probabilities between all hypotheses. In detail, two different strategies were evaluated:
(*i*) *Comparison after all optimization*. Firstly, a series of combinatorial optimization with all converters are performed. Then, the converter with the largest probability is chosen as the estimated anomaly (or normal).
(*ii*) *Comparison in situ*. The comparison is performed at the end of every cycle in the simulated annealing. In other words, the state "which converter is currently selected" is also dealt with as one extra variable to be optimized.

2.3. Evaluation

2.3.1. Virtual detector output construction. The detector outputs were virtually created for each landmark and for each CT dataset. If the target landmark existed in the subject's real body, it was included as a candidate, as well as 25, 50, 75 or 100 of false positive candidates. Each false positive candidate was determined randomly following a 3-D Gaussian probability distribution (the center corresponds to the true point, and the standard deviations were $\sigma_x = \sigma_y = \sigma_z = {}^{1}\!/_{\sqrt{3}} \cdot 100$ millimeters). Any false candidates within 20 millimeters from the true point were removed. If the target landmark did not exist in reality, only false positive points were added, using another adjacent landmark as the distribution center. The detector-estimated probability $p_m^{\exists}$ was fixed to be 0.05, and all of p_m^{i} were set to be uniform.
2.3.2. Experimental settings. The experiments were performed on a workstation with an Intel® Core™ i7-2600 processor and one NVIDIA® Tesla™ C2050 GPU

computing processor. The latter was utilized to speed up the calculation. The processing times were 9.7 minutes per case for *comparison after all optimization* strategy and 3.6 minutes for *comparison in situ* strategy with 100 candidates per landmark.

5 times of experiments were performed for each experimental setting, using different sets of virtual detector outputs.

The statistical model of squared distances (the mean $E(d_{i,j})$, variance $V(d_{i,j})$ and the covariance matrix **V**) was calculated by leave-one-case out method.

3 Results and discussion

The proposed algorithm was evaluated with 156 bony landmark points which were manually inputted in 50 thin-slice clinical human body CT datasets. The list of landmark used is available in [2]. Among the 50 datasets, total 9 had a segmentation anomaly (three 13T, one 11T, three 6L, two 4L and none of 13T+4L or 11T+6L).

The summary of results is shown in Table 1. An example result is also shown in Figure 4. The accuracies of anomaly estimation were from 87.6 to 97.6%, varying among strategies and number of candidates. The accuracy of anomaly estimation was better in *comparison after optimization* strategy than in *comparison in situ* strategy. Though the former strategy takes approximately 3 times longer time, it seems to be useful especially for a larger number of candidates.

Table 1. The entire result of anomaly estimation and combinatorial optimization.

	No. of cands	anomaly estimation accuracy (%)			landmark determination: existent-in-reality LMs (%)			non-existent LMs (%)	
		anomaly cases	normal cases	**overall**	TP	FN	$FP_{candidate}$	TN	$FP_{existence}$
comparison after opt.	25	88.9 ±13.6	99.5 ±1.1	**97.6 ±3.3**	96.60 ±2.90	0.09 ±0.13	3.31 ±2.78	100 ±0	0 ±0
	50	88.9 ±7.9	99.5 ±1.1	**97.6 ±1.7**	96.25 ±0.34	0.08 ±0.05	3.67 ±0.32		
	75	86.7 ±14.5	99.0 ±1.3	**96.8 ±2.7**	94.40 ±0.39	0.09 ±0.08	5.51 ±0.36		
	100	77.8 ±11.1	94.6 ±4.0	**91.6 ±3.0**	91.33 ±1.03	0.26 ±0.09	8.42 ±1.00		
comparison in situ	25	84.4 ±12.7	100 ±0	**97.2 ±2.3**	97.65 ±0.36	0.09 ±0.08	2.26 ±0.32		
	50	73.3 ±18.6	100 ±0	**95.2 ±3.4**	95.67 ±0.70	0.17 ±0.12	4.15 ±0.59		
	75	64.4 ±19.9	99.5 ±1.1	**93.2 ±3.9**	93.37 ±1.02	0.24 ±0.15	6.40 ±0.91		
	100	48.9 ±14.9	96.1 ±2.8	**87.6 ±2.6**	89.64 ±0.61	0.40 ±0.05	9.96 ±0.57		

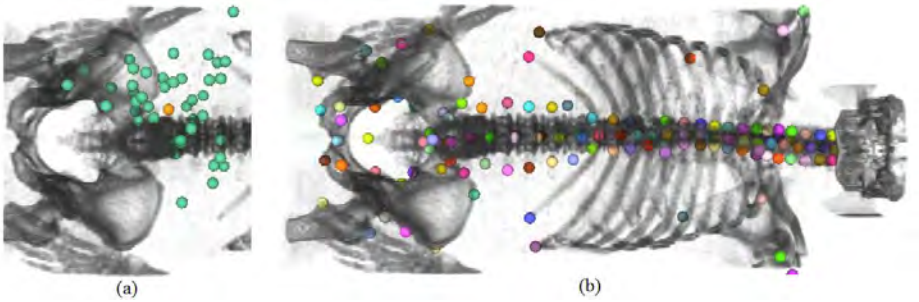


Fig. 4. An example result of a case with 6 lumbar vertebra (6L) anomaly. (a) The false (green) and true (orange) candidates outputted by the virtual detector for the tip of right transverse process of the 4th lumbar vertebra. (b) The optimization result for 156 landmarks.

The optimization result of each landmark in each case was also evaluated. Any landmark which existed in reality was classified as one of TP, $FP_{candidate}$ or FN. Within them, TP means that the algorithm correctly selected the true position candidate. $FP_{candidate}$ means that the algorithm selected any false position candidate. If the algorithm concluded that the target landmark did not exist, but in reality it existed, it was categorized as FN. In this study, the accuracy ratios (ratios of TP) for existent landmarks were varied from 89.64 to 97.65%.

Additionally, all of non-existing landmarks were classified as either TN or $FP_{existence}$. TN means that the target landmark did not exist in reality and the algorithm correctly found it (either by adopting some anomaly hypothesis which does not include the target landmark, or by selecting the state $\Im$ for the landmark.) If the algorithm chose any candidate other than $\Im$, it was classified as $FP_{existence}$. In this study all of non-existent landmarks were determined as TN, so the accuracy was 100%.

It is noticeable that, throughout all experimental condition, none of non-existing landmarks were mistaken as $FP_{existence}$. In fact, without any exception, all mistakes in anomaly estimation were underestimation of thoracic / lumbar vertebral number. It may imply that our algorithm was not sufficiently optimized for the problem, favoring anomalies having less vertebral number, and yet to be investigated in the future work. Also, we are now planning to overcome this limitation by adding some additional term, which evaluates the regularity of vertebral bones' alignment, to the MAP estimation.

4 Conclusion

A novel combinatorial optimization algorithm for landmark detection and anomaly estimation was presented. The proposed method showed fair results even with a large number of landmark position candidates. Additionally, its feasibility to estimate segmentation anomaly of the vertebrae, which is one of the most common and problematic anomalies in detecting bone landmarks, has been shown. Therefore, we believe that our algorithm is useful in medical image analysis such as a pre-

process for computer-assisted detection/diagnosis (CAD) applications. The future work will include improvement of accuracy in anomaly estimation and evaluation with real detector outputs with a large number of datasets.

Acknowledgement

This study is a part of the research project "Computational Anatomy for Computer-aided Diagnosis and therapy: Frontiers of Medical Image Sciences", which is financially supported by the grant-in-aid for scientific research on innovative areas MEXT, Japan.

References

1. Seifert, S., Barbu, A., Zhou, S.K., et. al.: Hierarchical parsing and semantic navigation of full body CT data. In: Pluim, J.P.W., Dawant, B.M. (eds.) SPIE (2009)
2. Nemoto, M., Masutani, Y., Hanaoka, S. et. al.: A unified framework for concurrent detection of anatomical landmarks for medical image understanding. In: Proc. SPIE 2011, 7962--121 (2011)
3. Crimi, A., Sporring, J., Bruijne, M., et. al.: Prior knowledge regularization in statistical medical image tasks. Probabilistic Models for Medical Image Analysis 2009, MICCAI, Imperial College London, UK (2009)
4. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.6, no.6, 721--741 (1984)
5. Carrino, J. A., Campbell, P. D., Lin, D.C., et al.: Effect of spinal segment variants on numbering vertebral levels at lumbar MR imaging. Radiology, 259:203--212 (2011)

Hippocampal Morphometry Study by Automated Surface Fluid Registration and its Application to Alzheimer's Disease

Jie Shi[1], Yuting Wang[2], Paul M. Thompson[3], and Yalin Wang[1]

[1] School of Computing, Informatics and Decision Systems Engineering, ASU, Tempe, AZ 85281, USA
[2] Mathematics Department, UCLA, Los Angeles, CA 90095, USA
[3] Lab. of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA 90095, USA
jie.shi@asu.edu

Abstract. Alzheimer's disease (AD) is a severe and growing public health crisis. Efforts are underway to look for AD early detection in an efficient manner. Among all the AD biomarkers, hippocampal atrophy assessed on high-resolution T1-weighted MRI is the best established and validated. Hippocampal morphometry is increasingly used in the AD research, with modeling the hippocampus as a 3D parametric surface mesh. However, a major question in the analysis is how to align corresponding surface regions across subjects. Here we develop a system for detecting AD symptoms on hippocampal surfaces with an automated surface fluid registration method, which is based on conformal surface representation and mutual information regularized image fluid registration. Since conformal mappings are diffeomorphic and the mutual information method is able to drive a diffeomorphic flow that is adjusted to enforce appropriate surface correspondences in the surface parameter domain, combining conformal and fluid mappings will generate 3D shape correspondences that are diffeomorphic. We also incorporate in the system a novel method to compute curvatures using surface conformal parameterization. Experimental results in three hippocampal datasets show that the new system outperformed an early similar method and the popular SPHARM tool.

1 Introduction

Alzheimer's disease (AD) doubles in frequency of onset every 5 years after age 60, afflicting 1% of those aged 60 to 64, and 30-40% of those over 85. Many MRI-based measures of atrophy in several structural measures, including whole-brain, hippocampus [14, 16, 19], and ventricular enlargement [19], etc., correlate closely with changes in cognitive performance, supporting their validity as markers of AD progression [8]. Among all the biomarkers, hippocampal atrophy assessed on high-resolution T1-weighted MRI is the best established and validated. As a result, detection of valid and efficient morphometry changes in hippocampus and their correlation with other cognitive functions and biomarkers becomes a

key research topic for clinical diagnosis and monitoring of patients with suspected Alzheimer's disease. Although most hippocampus analysis used volume as the atrophy measurement [10], recent researches [1, 16, 19, 26] demonstrated that surface-based subcortical structure analysis may offer more advantages because these methods studied patterns of hippocampal subfield atrophy and produced detailed point-wise correlation between atrophy and cognitive functions/biological markers [19].

Brain surface deformation studies typically require the computation of dense correspondence vector fields that match one surface with another. Many brain surface registration methods have been proposed [6, 20, 27].Often, higher order correspondences must be enforced between specific anatomical points and curved landmarks lying within the two surfaces. This can be achieved by first mapping each of the 3D surfaces to canonical parameter spaces such as a sphere [7] or a planar domain [21]. A flow, computed in the parameter space of the two surfaces [4], induces a correspondence field in 3D. Artificial neural networks can rule out or favor certain types of feature matches [15]. Correspondences may be determined by using a minimum description length (MDL) principle, based on the compactness of the covariance of the resulting shape model [5]. A key direction in surface registration research has been the computation of a diffeomorphic surface map that also matches automatically identified surface features.

Using holomorphic 1-forms, a global conformal parameterization can be developed to conformally map a surface with complex topology (e.g., a surface with branching topology) to a set of rectangular domains in the Euclidean plane. The resulting parameterization helps in discretizing partial differential equations (PDEs) for smoothing, denoising, or matching signals defined on the surface. The mutual information (MI) method has been widely used to drive a diffeomorphic flow in image registration. By adjusting the mutual information method to enforce appropriate surface correspondences in the parameter domain, any scalar-valued signals defined on the surfaces can also be aligned using the same flow field. Conformal maps and fluid registration techniques can be combined to avoid having to define a large set of manually-defined landmarks to constrain brain surface correspondences. Since they generate diffeomorphic mappings, conformal and fluid mappings together could generate 3D shape correspondences that are diffeomorphic (i.e., smooth one-to-one correspondences). In [21, 22], Wang et al. proposed an automated surface fluid registration method based on conformal mapping and mutual information regularized image fluid registration and applied it to register human faces and hippocampus. Here we develop a system based on this technique for studying hippocampus in AD and incorporate a novel method to compute surface curvatures as proposed in [12]. Our major contributions can be summarized as: (1). Introduction of a new stable method to compute surface curvatures. (2). An automated hippocampal surface registration system validated in three AD datasets with better performance than a previous similar method [23] and SPHARM [18]. (3). The system will be publicly available [24]. Last, although the current system finds applications in AD detection, it is a general method which may be applied to many other applications.

2 Method

2.1 Surface Conformal Parameterization

Let S be a surface in $\mathbb{R}^3$ with an atlas $\{(U_\alpha, z_\alpha)\}$, where (U_α, z_α) is a coordinate chart defined on S. The atlas thus is a set of consistent charts with smooth transition functions between overlapping charts. Here $z_\alpha : U_\alpha \to \mathbb{C}$ maps an open set $U_\alpha \subset S$ to a complex plane $\mathbb{C}$. If on any chart (U_α, z_α) in the atlas, the Riemannian metric or the first fundamental form can be formulated as $ds^2 = \lambda(z_\alpha)^2 dz_\alpha d\bar{z_\alpha}$, and the transition maps $z_\beta \circ z_\alpha^{-1} : z_\alpha(U_\alpha \bigcap U_\beta) \to z_\beta(U_\alpha \bigcap U_\beta)$ are holomorphic, the atlas could be called conformal. Given a conformal atlas, a chart is compatible with the atlas if adding this chart still generates a conformal atlas. A conformal structure is obtained by adding all possible compatible charts to a conformal atlas. A Riemann surface is a surface with a conformal structure. One coordinate chart in the conformal structure introduces a *conformal parameterization* between a surface patch and the image plane. The conformal parameterization is angle-preserving and intrinsic to the surface geometry.

Let (x_α, y_α) be the local parameter on the chart (U_α, z_α) on S, a differential 1-form in (x_α, y_α) can be defined as

$$\omega = f(x_\alpha, y_\alpha)dx_\alpha + g(x_\alpha, y_\alpha)dy_\alpha \tag{1}$$

where f, g are smooth functions. ω is a closed 1-form, if in each local parameter (x_α, y_α), $\frac{\partial f}{\partial y_\alpha} - \frac{\partial g}{\partial x_\alpha} = 0$. If ω is the gradient of another function defined on S, it can be called an exact 1-form. An exact 1-form is also a closed 1-form. If a closed 1-form ω satisfies $\frac{\partial f}{\partial x_\alpha} + \frac{\partial g}{\partial y_\alpha} = 0$, then it is a harmonic 1-form. The gradient of a harmonic 1-form is an exact harmonic 1-form. The Hodge star operator acting on a differential 1-form gives the conjugate differential 1-form

$$^*\omega = -g(x_\alpha, y_\alpha)dx_\alpha + f(x_\alpha, y_\alpha)dy_\alpha \tag{2}$$

Intuitively, the conjugate 1-form $^*\omega$ is obtained by rotating ω by a right angle everywhere. If ω is harmonic, so is its conjugate $^*\omega$. The holomorphic 1-form consists of a pair of conjugate harmonic 1-forms:

$$\tau = \omega + \sqrt{-1}^*\omega \tag{3}$$

For a Riemann surface S with genus $g > 0$, its conformal structure can always be represented in terms of a holomorphic 1-form basis, which is a set of $2g$ functions $\tau_i : K_i \to R^2, i = 1, 2 \cdots, 2g$ [22]. Any holomorphic 1-form τ is a linear combination of these functions. This finite-dimensional linear space generates all possible conformal parameterizations of surface S and the quality of a global conformal parameterization is fundamentally determined by the choice of the holomorphic 1-form [22]. Then the conformal parameterization ϕ at point p can be computed by integrating the holomorphic 1-form: $\phi(p) = \int_\gamma \tau$, where γ is any path joining p to a fixed point c on the surface. Figure 1 (a) illustrates a brain cortical surface and its conformal parameterization to a square.

2.2 Surface Conformal Representation

It has been known that surface registration requires defining a lot of landmarks in order to align corresponding functional regions. Labeling features could be accurate but time-consuming. Here we show that surface conformal parameterization could represent surface geometric features, thus avoiding the manual definition of landmarks.

For a general surface and its conformal parameterization $\phi : S \to \mathbb{R}^2$, the conformal factor at a point p can be determined by the formulation:

$$\lambda(p) = \frac{\text{Area}(B_\epsilon(p))}{\text{Area}(\phi(B_\epsilon(p)))} \tag{4}$$

where $B_\epsilon(p)$ is an open ball around p with a radius ϵ. The conformal factor λ encodes a lot of geometric information about the surface and can be used to compute curvatures and geodesic. In our system, we compute the surface mean curvatures only from the derivatives of the conformal parameterization as proposed in [12], instead of the three coordinate functions and the normal, which are generally more sensitive to digitization errors. Mathematically, the mean curvature is defined as:

$$H = \frac{1}{2\lambda}\text{sign}(\phi)|\Delta\phi|, \text{ where } \text{sign}(\phi) = \frac{< \Delta\phi, \overrightarrow{N} >}{|\Delta\phi|}. \tag{5}$$

Using this formulation of H, we need to use the surface normal $\overrightarrow{N}$ only when computing sign(ϕ), which takes the value 1 or -1. Thus, the surface normal does not need to be accurately estimated and still we can get more accurate mean curvatures. Using the Gauss and Codazzi equations, one can prove that the conformal factor and mean curvature uniquely determine a closed surface in $\mathbb{R}^3$, up to a rigid motion. We call them the *conformal representation* of the surface. Figure 1 (b) shows the computed conformal factor (left) and mean curvature (right) on a hippocampal surface with color indices according to the values. Since conformal factor and mean curvature could represent important surface features and they are intrinsic to the surface, they may be used for surface registration.

2.3 Surface Fluid Registration Regularized by Mutual Information

After computing intrinsic geometric features, we align surfaces in the parameter domain with a fluid registration technique.Using conformal mapping, we essentially convert the surface registration problem to an image registration problem. The mutual information (MI) method has been successfully used to drive a diffeomorphic flow in rigid [25] and non-rigid [13, 17] image registration. Image registration will be optimized when MI between two images is maximized. For MI to work, a monotonic mapping in grayscales between images is not required, so images from different modalities can be registered [11]. Hermosillo et al. [9] adopted linear elasticity theory to regularize the variational maximization of MI.

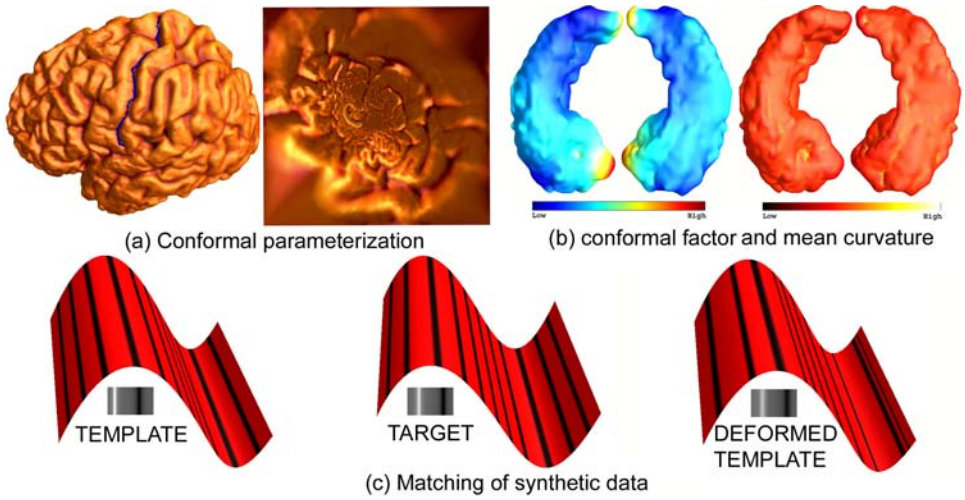


Fig. 1. Illustration of conformal parameterization (a) and geometric features (b). (c) shows the matching of geometric features in the 2D parameter domains using fluid registration with synthetic surfaces. Geometric features on 3D surfaces were computed and mapped to 2D conformal parameter domains.

D'Agostino et al. [3] extended this approach to a viscous fluid scheme allowing large local deformations, while maintaining smooth, one-to-one topology [2]. We call this approach MI regularized fluid registration.

In [21, 22], Wang et al. proposed an automated surface fluid registration method combining conformal mapping and image fluid registration [3]. Let I_1, I_2 be the conformal representations of the target and the deforming template surfaces, respectively, the MI between two surfaces was defined as [21, 22]

$$I(u) = \int_{\mathbb{R}^2} p_u(i_1, i_2) log \frac{p_u(i_1, i_2)}{p(i_1)p_u(i_2)} di_1 di_2 \tag{6}$$

where $p(i_1) = P(I_1(x) = i_1), p_u(i_2) = P(I_2(x-u) = i_2)$, and $p_u(i_1, i_2) = P(I_1(x) = i_1 \ \& \ I_2(x-u) = i_2)$. Since conformal mapping and MI regularized fluid registration generate diffeomorphic mappings, a diffeomorphic surface-to-surface mapping is then recovered that matches surfaces in 3D. In our system, we adopt their methods of conformal mapping and fluid registration. However, our system differs from theirs in the computation of surface features as introduced in Sec. 2.2. The new way to compute mean curvature is more stable and less sensitive to normal computation, thus gives better representation of the surface features for registration.

3 Results

3.1 Synthetic Surface Registration Result

First, we illustrate the algorithm on synthetic surfaces. Figure 1 (c) illustrates a synthetic surface example. A pair of simple S-shape surfaces were generated. Corresponding 2D images were generated based on the sum of the local conformal factor and the mean curvature, expressed in the conformal parameterization domain. Some black horizontal lines were drawn on the surfaces to show equal distances on the surfaces and represent the differences in their shapes.The locations of the highest and lowest intensities are different (as shown by the positions of the horizontal stripes in the 2D images below). Using surface-based fluid registration, in the last image, the obtained horizontal line positions demonstrated an improved matching between features lying in the two surfaces.

In this experiment, we aim to visually explain the work flow of the algorithm. First, we conformally mapped each surface onto a planar domain and computed their conformal representation by combining conformal factor and mean curvature. Second, we scaled the conformal representation to form the 2D feature images as shown by the first two 2D images in Fig. 1 (c). As we pointed out in Sec. 2.2, the 2D images, i.e. the conformal representations of the surfaces clearly show the characteristics of the surfaces. Third, by registering the two feature images using fluid registration method, we induced a change in both the template feature image and the template surface. The last image in Fig. 1 (c) demonstrates that without changing the shape of the surface, the features on the template surface are well aligned with the target surface.

3.2 Hippocampal Surface Morphometry in Alzheimer's Disease

In this experiment, we test the robustness of our system by applying it to two clinical studies of hippocampal changes in Alzheimer's disease (AD). In our system, we leave two holes at the front and back of the hippocampal surface, representing its anterior junction with the amygdala, and its posterior limit as it turns into the white matter of the fornix. The resulting structure can then be logically represented as an open boundary genus-one surface, i.e., a cylinder. To better visualize the matching of surface features, we chose to encode surface features using a compound scalar function based on the local conformal factor and the mean curvature. A similar technique for AD study in hippocampus was proposed in [23], but with surface registration using constrained harmonic map. We take their method as a comparison. After the cross-subject registration is computed with one target surface selected, we examine shape differences using the multivariate tensor-based morphometry (MTBM) [23]. MTBM computes statistics from the Riemannian metric tensors that retain the full information in the deformation tensor fields, thus is more powerful in detecting surface differences than many other statistics [23].

Figure 2 illustrates the experimental results on a group of hippocampal surface models extracted from 3D brain MRI scans of 12 individuals with AD and

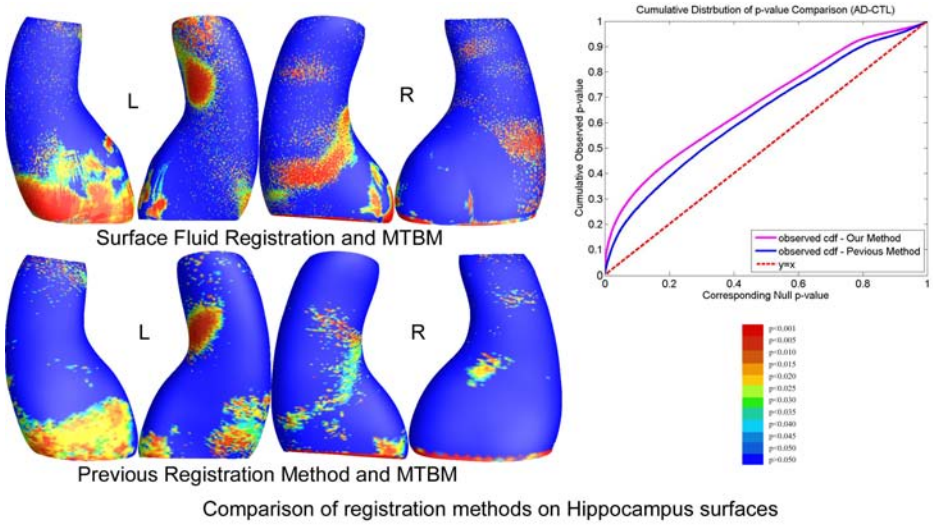


Fig. 2. The comparison of our method and a previous similar method on map of local shape differences (p-values), based on the multivariate TBM method with hippocampal surfaces in 12 AD versus 14 control subjects [19].

14 matched healthy control subjects, the same dataset as in [19], where the data was manually traced by experienced neurologists. With MTBM, we can see that the significant areas detected with two methods are consistent.But with surface fluid registration, we detected larger areas with more significant group differences in the surface parameterization tensor, which is related to the relative area of regions in disease versus normality.The overall statistical significance of these group difference maps, based on permutation testing of the suprathreshold area of statistics (and therefore corrected for multiple comparisons) were p=0.0063 for the left hippocampal surface and 0.0298 for the right. This outperformed the previous method [23] (0.0205 for the left hippocampal surface and 0.1026 for the right). The cumulative distribution function (CDF) plot shows the comparison of resulting p-values for each method. While the line $y = x$ represents null hypothesis, which is according to no difference, steeper curve shows more differences detected.

Figure 3 illustrates the experimental results on another hippocampal dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (http://www.loni.ucla.edu/ADNI). Mild cognitive impairment (MCI) is an intermediate stage between the expected cognitive decline of normal aging and the more pronounced decline of dementia. If MCI could be found and treated, the risk of AD will be significantly reduced. However, at MCI stage, changes in brain surface are not significant thus impose more difficulty on the detection. Here we randomly chose 40 AD, 40 MCI, and 40 control subjects to test the effectiveness of our system. The hippocampus data was automatically segmented by a prior work [14]. With MTBM, we can see that, in the three experiments, our system

demonstrated better results than the previous method. Particularly, our system gave better MCI detection results when comparing with both AD and control subjects. In the experiment, all group difference p-maps were corrected using false discovery rate (FDR). The FDR method decides whether a threshold can be assigned to the statistical map that keeps the expected FDR below 5% (i.e., no more than 5% of the voxels are false positive findings). The CDF plots show the uncorrected p-values (as in a conventional FDR analysis). The x value at which the CDF plot intersects the $y = 20x$ line represents the FDR corrected p-value or q-value. It is the highest statistical threshold that can be applied to the data, for which at most 5% false positives are expected in the map. In general, a larger q-value indicates a more significant difference in the sense that there is a broader range of statistic threshold that can be used to limit the rate of false positives to at most 5%. The use of the $y = 20x$ line is related to the fact that significance is declared when the volume of suprathreshold statistics is more than 20 times that expected under the null hypothesis. Table 1 gives the FDR corrected p-values comparison.

3.3 Comparison with SPHARM

In this study, we tested our system on the whole ADNI baseline dataset (http://www.loni.ucla.edu/ADNI). The dataset consists of 233 healthy controls, 410 subjects with MCI, and 200 patients with AD. We excluded 1 subject from the control group and 2 subjects from the MCI group due to name duplication. For subjects with duplicated names, we retained the one which is the repeated scan. The hippocampal surfaces were automatically segmented using FIRST (http://www.fmrib.ox.ac.uk/fsl/first/index.html). FIRST is an integrated surface analysis tool developed as part of the FSL library, which is written mainly by members of the Analysis Group, FMRIB, Oxford, UK. We compared our system's results with the popular SPHARM tool [18] in surface registration on the same dataset. We adopted the suggested parameters for hippocampus in the manual of SPHARM [18]. In this experiment, 1 subject from each group (AD, MCI, control) failed the FIRST segmentation step probably due to the original images' resolution or contrast. After FIRST segmentation, we extracted the left and right hippocampi and saved each of them into a binary image. We fed the binary images as the input of both systems. As a result, 231 control, 199 AD, and 407 MCI subjects were successfully registered by our system, which shows the robustness of the system compared with SPHARM, in which 5 control, 17 AD, 14 MCI subjects failed either due to segmentation failure or parameterization failure. Fig. 4 shows the experimental results. From the p-map and the CDF plots, we can see that, with MTBM, our system outperformed the SPHARM

	Our Method	Prior Method [23]
AD-CTL	0.0476	0.0467
MCI-AD	0.0217	0.0199
MCI-CTL	0.0221	0.0185

Table 1. FDR corrected p-values on hippocampal surfaces

in all three group difference studies. Table 2 gives the FDR corrected p-values comparison. Considering fairness, we also made comparisons by excluding those subjects that failed in SPHARM from our system in the statistical study. Table 3 gives the FDR corrected p-values comparison with the failed subjects excluded, i.e., 226 control, 182 AD, and 393 MCI subjects were studied in this experiment. From the table, we can see that our system still outperformed SPHARM on this dataset. Furthermore, because AD and MCI failed more than control subjects, the AD-MCI group has been more affected as expected.

4 Conclusion and Future Work

We develop an automated surface fluid registration system for hippocampal surface registration. Experiments on different types of AD hippocampal datasets all demonstrate our system's stronger statistical power. Ongoing work is to apply this system to automatically map lateral ventricle enlargements in Alzheimer's disease and those at risk.

References

1. Apostolova, L.G., et al.: Subregional hippocampal atrophy predicts Alzheimer's dementia in the cognitively normal. Neurobiol Aging 31(7), 1077–1088 (2008)
2. Christensen, G.E., et al.: Deformable templates using large deformation kinematics. IEEE Trans Image Process 5(10), 1435–1447 (1996)
3. D'Agostino, E., et al.: A viscous fluid model for multimodal non-rigid image registration using mutual information. Med Image Anal 7(4), 565–575 (2003)
4. Davatzikos, C.: Spatial normalization of 3D brain images using deformable models. J. Comp. Assisted Tomography 20(4), 656–665 (1996)
5. Davies, R.H., et al.: A minimum description length approach to statistical shape modeling. IEEE Trans Med Imaging 21(5), 525–537 (2002)
6. Durrleman, S., et al.: Sparse approximation of currents for statistics on curves and surfaces. In: Med. Image Comp. Comput.-Assist. Intervention, Proceedings. vol. 11, pp. 390–398 (2008)
7. Fischl, B., et al.: High-resolution inter-subject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8(4), 272–284 (1999)
8. Frisoni, G.B., et al.: The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology 6(2), 67–77 (2010)
9. Hermosillo, G.: Ph.D. thesis, Université de Nice (INRIAROBOTVIS), Sophia Antipolis, France (2002)
10. Holland, D., et al.: Subregional neuroanatomical change as a biomarker for Alzheimer's disease. In: Proc. Natl. Acad. Sci. USA (2009)
11. Kim, B., et al.: Mutual information for automated unwarping of rat brain autoradiographs. Neuroimage 5(1), 31–40 (1997)

	Our Method	SPHARM [18]
AD-CTL	0.0485	0.0459
MCI-AD	0.0215	0.0047
MCI-CTL	0.0483	0.0288

Table 2. FDR corrected p-values on hippocampal surfaces.

12. Lui, L.M., et al.: Computation of curvatures using conformal parameterization. Communications in Information and Systems 8(1), 1–16 (2008)
13. Meyer, C.R., et al.: Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. Medical Image Analysis 1(3), 195–206 (1997)
14. Morra, J.H., et al.: Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Neuroimage 45(1 Suppl), S3–15 (2009)
15. Pitiot, A., et al.: Learning object correspondences with the observed transport shape measure. In: Proceedings of Information Processing in Medical Imaging (IPMI 2003). pp. 25–37 (2003)
16. Qiu, A., et al.: Regional shape abnormalities in mild cognitive impairment and Alzheimer's disease. Neuroimage 45(3), 656–661 (2009)
17. Rueckert, D., et al.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging 18(8), 712–721 (1999)
18. Styner, M., et al: Framework for the statistical shape analysis of brain structures using SPHARM-PDM. The Insight Journal pp. 242–250 (2006)
19. Thompson, P.M., et al.: Mapping hippocampal and ventricular change in Alzheimer disease. Neuroimage 22(4), 1754–1766 (2004)
20. Vaillant, M., et al.: Surface matching via currents. In: Proceedings of Information Processing in Medical Imaging (IPMI 2005). vol. 19, pp. 381–392 (2005)
21. Wang, Y., et al.: Automated surface matching using mutual information applied to Riemann surface structures. In: Med. Image Comp. Comput.-Assist. Intervention, Proceedings, Part II. pp. 666–674 (2005)
22. Wang, Y., et al.: Mutual information-based 3D surface matching with applications to face recognition and brain mapping. In: Proc. Intl Conf. Computer Vision. pp. 527–534 (2005)
23. Wang, Y., et al.: Multivariate tensor-based brain anatomical surface morphometry via holomorphic one-forms. In: Med. Image Comp. Comput.-Assist. Intervention, Proceedings. vol. 12, pp. 337–344 (2009)
24. Wang, Y.: Multivariate tensor-based subcortical morphometry system (2011), http://gsl.lab.asu.edu/conformal.htm
25. West, J., et al.: Comparison and evaluation of retrospective intermodality brain image registration techniques. J. Comp. Assisted Tomography 21(4), 554–566 (1997)
26. Yushkevich, P.A.: Continuous medial representation of brain structures using the biharmonic PDE. Neuroimage 45(1 Suppl), S99–110 (2009)
27. Zhong, J., et al.: Multi-manifold diffeomorphic metric mapping for aligning cortical hemispheric surfaces. Neuroimage 49(1), 355–365 (2010)

	Our Method	SPHARM [18]
AD-CTL	0.0486	0.0459
MCI-AD	0.0195	0.0047
MCI-CTL	0.0483	0.0288

Table 3. FDR corrected p-values on hippocampal surfaces (with failed subjects excluded).

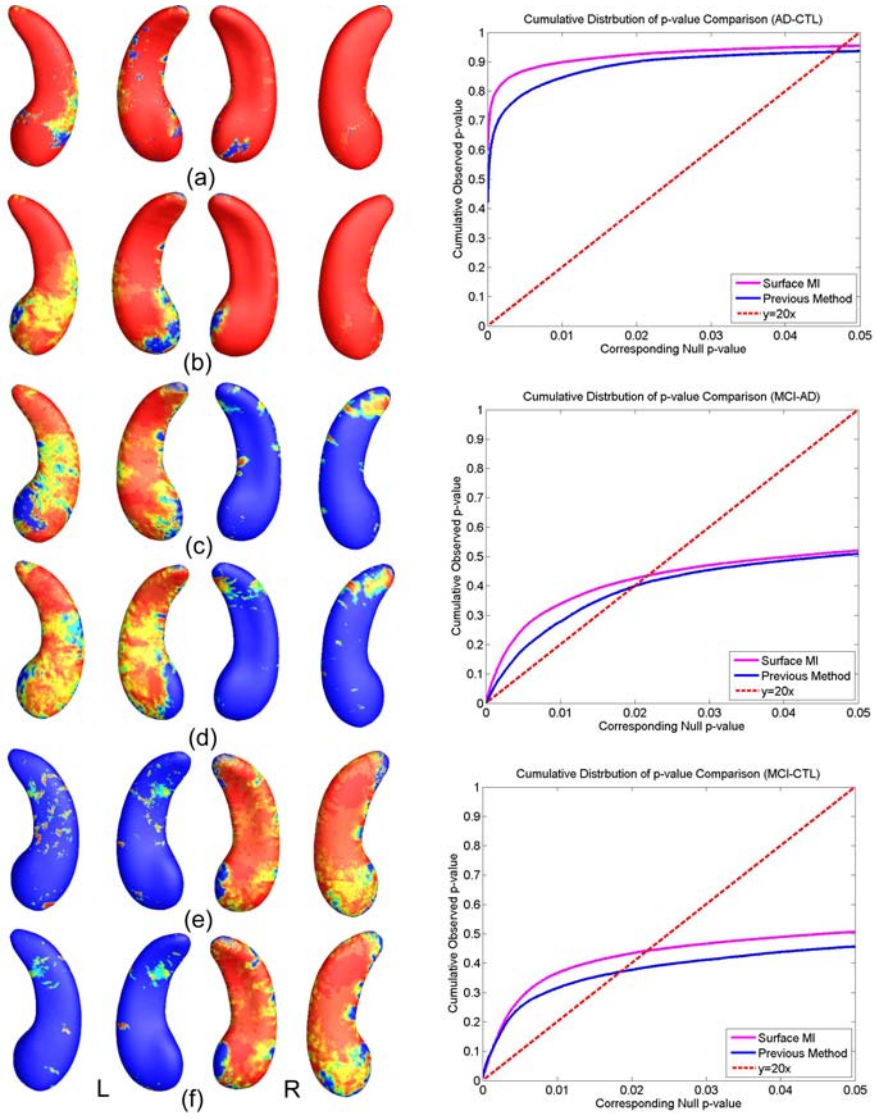


Fig. 3. Comparison of surface fluid registration and the previous method on map of local shape differences (p-values), based on the mutivariate TBM method with hippocampal surfaces in 40 AD, 40 MCI, and 40 control subjects which were automatically segmented [14]. (a), (c), (e) are our results on group difference between AD and control, MCI and AD, MCI and control, respectively. Similarly, (b), (d), (f) are results of a prior work [23] on AD and control, MCI and AD, MCI and control, respectively. The p-map color scale is the same as Figure 2.

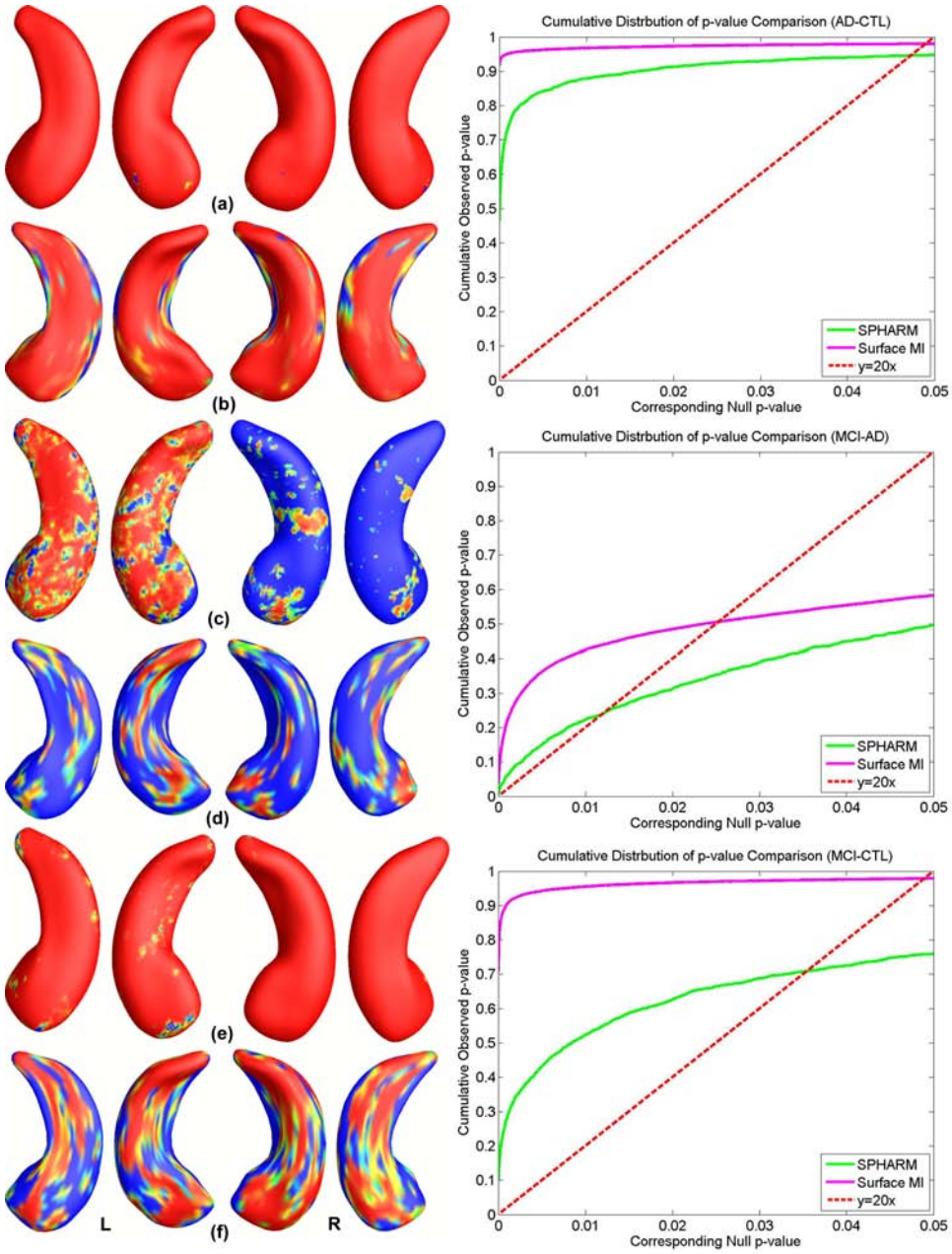


Fig. 4. Comparison of surface fluid registration and SPHARM on map of local shape differences (p-values), based on the mutivariate TBM method with hippocampal surfaces from ADNI baseline data, which were automatically segmented by FIRST. (a), (c), (e) are our results on group difference between AD and control, MCI and AD, MCI and control, respectively, in 231 control, 199 AD, and 407 MCI subjects. Similarly, (b), (d), (f) are results of SPHARM on AD and control, MCI and AD, MCI and control, respectively, in 226 control, 182 AD, and 393 MCI subjects. The p-map color scale is the same as Figure 2.

A New Riemannian Setting for Surface Registration

Martin Bauer[1], Martins Bruveris[2]

[1] Fakultät für Mathematik, Universität Wien, Nordbergstrasse 15, A-1090 Wien
[2] Dep. of Mathematics, Imperial College, London SW7 2AZ, UK

Abstract. We present a new approach for matching regular surfaces in a Riemannian setting. We use a Sobolev type metric on deformation vector fields which form the tangent bundle to the space of surfaces. In this article we compare our approach with the diffeomorphic matching framework. In the latter approach a deformation is prescribed on the ambient space, which then drags along an embedded surface. In contrast our metric is defined directly on the deformation vector field and can therefore be called an *inner metric*. We also show how to discretize the corresponding geodesic equation and compute the gradient of the cost functional using finite elements.

Keywords: Registration, Surface Matching, LDDMM, Computational Anatomy, Geodesic Shooting, Adjoint Equations

1 Introduction

The field of computational anatomy concerns itself with the study and classification of the variability of biological shapes, including their statistical variance. Since the space of all shapes is inherently nonlinear, the usual methods of linear statistics cannot be applied. In particular, the addition of two surfaces cannot be meaningfully defined. One way to overcome this difficulty is to introduce a Riemannian structure on the space of shapes, which locally linearizes the space and allows the development of statistical methods that are analogous to the linear case. This approach was taken, e.g., in [7]. In the Riemannian setting, the average of two shapes may be defined as the middle point of a geodesic joining these two shapes. In a similar way one may define the corresponding geodesic mean of a collection of n shapes.

One class of shapes which are of interest in computational anatomy consists the surfaces embedded in $\mathbb{R}^3$. The cortical surface, the surfaces of hippocampi, thalami, and nasal cavities are all examples of shapes which are represented as two dimensional surfaces in $\mathbb{R}^3$. This is also an example, where the Riemannian setting may be applied to study collections of shapes.

Throughout the last decade various (Riemannian) metrics have been proposed. They include a metric that preservers local rigidity [11], a generalization of the elastic metric for curves to higher dimensions [18, 13], a metric inspired by a continuum mechanics which is defined in the interior of a two dimensional

shape [23, 8] and a metric based on the square-root representation of surfaces [12]. Other approaches include the representation of surfaces via densities [20] and metrics defined on surfaces via level sets [15, 19].

Another method for comparing anatomical shapes in the Riemannian setting is the method of large deformation diffeomorphic metric matching (LDDMM), based on the deformable template paradigm of Grenander [10]. In this setting, a template shape is matched to a target shape by finding a transformation in a suitable group of deformations of the ambient space that transforms the template into the target. This approach has been systematically developed in [3, 4, 9, 21, 22] and applied to various problems in computational anatomy. Registering two surfaces in this framework involves finding a diffeomorphism of the whole space, which transforms one surface to the other. Because of its widespread use in the field of computational anatomy for registering volumetric images we will use LDDMM as a reference to highlight the features of our proposed framework.

In this paper we propose a different way of defining a Riemannian structure on the space of surfaces, which also provides the full range of tools for nonlinear statistics. We use a Sobolev type norm to enforce regularity of the deformation vector field and measure the cost of the deformation. Our approach to characterizing a deformation is intrinsic to the surface, rather than resulting from a transformation of the surface induced by a deformation of the ambient space in which the image is embedded. For this reason, the Riemannian metrics used here are called *inner metrics* as opposed to the *outer metrics* used in LDDMM, where deformations are imposed via the ambient space. Other examples for inner metrics can be found in [18, 12, 23, 8, 1].

Inner metrics of Sobolev type on planar curves were introduced and studied previously in [17, 24]. Recently they were generalized to surfaces and higher dimensional hypersurfaces in Euclidean space in [2]. The numerical implementation of matching with these metrics differs from LDDMM, because the metric on the tangent space at each surface depends nonlinearly on the surface. This means the metric will change adaptively as one moves around in shape space. This adaptive property is in marked contrast to LDDMM, where the metric is defined on the diffeomorphisms and projected down to the shapes, so it doesn't depend on the particular shape.

The outline of the paper is as follows. In Sec. 2 we review the registration problem for surfaces and recall how it is solved using outer metrics in LDDMM. Then we present the approach via inner metrics of Sobolev type and point out the differences between the two methods. For definiteness, we will concentrate our attention on the Sobolev metric of order one. In Sec. 3 we discuss how to discretize and implement the geodesic equations for this metric and how to solve the registration problem via geodesic shooting. Finally, in Sec. 4 we show how this metric performs in some examples using synthetic data.

2 The Mathematical Formulation

We are dealing with the registration of parametrized regular surfaces. Such a surface is given by a smooth function $q : M \to \mathbb{R}^3$ from a model surface M into the Euclidean space $\mathbb{R}^3$. We will consider different choices of the model surface M in this paper: the plane sheet $M = [0,1] \times [0,1]$, the cylinder $M = S^1 \times [0,1]$ and the torus $M = S^1 \times S^1$. Another interesting choice would be the sphere $M = S^2$, which however is not considered in this paper. The metric can be defined in the same way as for the other topologies, however the numerical treatment is more challenging, because the sphere cannot be covered by a single global coordinate chart. We require the parametrization of the surface q to be regular in the following sense: at each point $x \in M$ the partial derivatives $\frac{\partial q}{\partial x^1}$, $\frac{\partial q}{\partial x^2}$ are required to be linearly independent. We will denote the space of all such surfaces by $\mathscr{S}$.

2.1 Registration with LDDMM

In the LDDMM framework, the registration of a template surface q_0 to a target surface q_{targ} involves finding a curve φ_t of diffeomorphisms of the ambient space $\mathbb{R}^3$, whose deformation carries the template surface to the target surface. Mathematically, one constructs these deformations using time dependent vector fields $v_t(y)$, which generate φ_t as their flow, i.e.

$$\partial_t \varphi_t = v_t(\varphi_t) \ . \tag{1}$$

The registration problem consists of finding a vector field, which minimizes the following sum of a kinetic energy and a matching term

$$E(v_t) = \frac{1}{2} \int_0^1 \|v_t\|_V^2 \, dt + \frac{1}{2\sigma^2} d(\varphi_1(q_0), q_{\text{targ}}) \ . \tag{2}$$

The kinetic energy is usually measured using a norm $\|.\|_V$ defined on a reproducing kernel Hilbert space V of vector fields on $\mathbb{R}^3$ with kernel K. The norm is then given by $\|u\|_V^2 = \int u \cdot K^{-1} \star u \, dx$. We will discuss possible choices of the matching term in Section 2.5.

It is possible to reduce the complexity of the problem, since one can show that minimizing vector fields v_t have to obey an evolution equation. This enables us to describe the whole vector field v_t by knowing only its value v_0 at time $t = 0$. The equations

$$\partial_t q_t(x) = v_t(q_t(x)) \tag{3a}$$

$$\partial_t p_t(x) = (Dv_t(q_t(x)))^T p_t(x) \tag{3b}$$

$$v_t(y) = \int_M p_t(x) K(y - q_t(x)) \, dx \tag{3c}$$

are given in terms of a momentum p_t, which lives on the surface. The momentum is convolved with the kernel in (3c) to reconstruct the minimizing vector field,

which defines the deformation of the ambient and drags the surface along in (3a). Details about LDDMM and the reformulation as evolution equations can be found in [3, 25].

2.2 Registration with Inner Metrics

We propose to use a different approach, described from a mathematical point of view in [1, 2]. In this approach we describe the deformation of the surface directly, without assuming an underlying deformation of the whole space. In our approach we will replace (3a) by

$$\partial_t q_t(x) = u_t(x) \ , \tag{4}$$

where $u_t(x) \in C^\infty(M, \mathbb{R}^3)$ is a time dependent vector field, defined only on the surface. Note the difference between the vector field $v_t(y)$ which is defined on $\mathbb{R}^3$ and $u_t(x)$, which is defined on the model space M, c.f. Fig. 1.

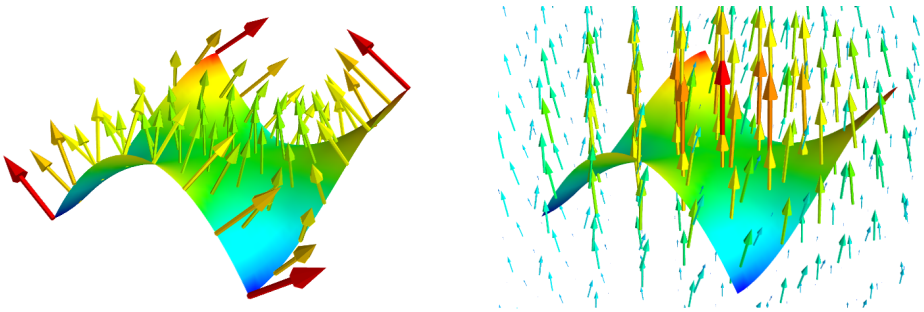


Fig. 1. For inner metrics the vector field governing the deformation is defined directly on the surface (left picture). In contrast the LDDMM approach defines a deformation vector field on all of $\mathbb{R}^3$ (right picture). The latter vector field deforms the ambient space and in the process induces a deformation of the surface.

In the new framework the registration problem still consists of minimizing the following sum of a kinetic energy term and a matching functional

$$E(u_t) = \frac{1}{2}\int_0^1 \langle u_t, u_t\rangle_{q_t}\, dt + \frac{1}{2\sigma^2} d(q_1, q_{\text{targ}}) \ . \tag{5}$$

The kinetic energy is measured via an inner product $\langle ., . \rangle_{q_t}$ on the space of vector fields along the surface. This inner product can and usually will depend nonlinearly on the surface q_t itself. This is another difference with the LDDMM framework, where the kinetic energy of the vector fields didn't depend on the surfaces that were matched. There is a whole variety of inner metrics, one can choose from. We will concentrate in this paper on an H^1-type metric, which will be discussed in Sec.2.4. Again we postpone the discussion of the matching term to Sec. 2.5.

2.3 Geometry

Both the LDDMM approach and the inner metrics can be seen as a special case of a constructions in the general framework of Riemannian geometry. In the case of inner metrics the collection of inner products $\langle .,. \rangle_q$ defines a Riemannian metric on the space $\mathscr{S}$ of all surfaces. In Riemannian geometry, curves, which minimize the energy $\int_0^1 \langle u_t, u_t \rangle_{q_t} \, dt$ for fixed endpoints q_0, q_1 are called geodesics. We can see that minimizers of (5) have to be geodesics in the space $\mathscr{S}$.

In LDDMM the inner product on the space of vector fields V also defines a Riemannian metric, this time on the group of diffeomorphisms. Therefore the minima of the registration problem (2) generates geodesics in the diffeomorphism group.

Why is it advantageous to work in a Riemannian setting? In this setting, the minima of the matching energy are geodesics, so one may describe the nonlinear space of shapes $\mathscr{S}$ in terms of the initial velocity or momentum of a geodesic, which is an element in a linear vector space. This is possible, because geodesics obey an evolution equation like (3). Using the initial velocity, which encodes the whole solution of the registration problem, we are able to view the space of surfaces from the template surface q_0 as a linear space. This enables us to use statistics, compute average surfaces and measure distances.

2.4 H^1-type Metric on Surfaces

We will consider a metric on the space of surfaces, which is the analogue of the H^1-norm $\|f\|^2_{H^1} = \int_{\mathbb{R}^2} |f(x)|^2 + |\nabla f(x)|^2 \, dx$ for functions on $\mathbb{R}^2$. We will replace functions on $\mathbb{R}^2$ by vector fields living on the curved surface q and adjust the definition of the H^1-norm to take into account the curved nature of the surface q. Since q is a surface in $\mathbb{R}^3$, we can measure angles and distances of vectors tangent to the surface, using the Euclidean inner product on $\mathbb{R}^3$. At each point $q(x)$ of the surface we also have a canonical basis for the plane tangent to q, given by the vectors $\frac{\partial q}{\partial x^1}$, $\frac{\partial q}{\partial x^2}$. We denote the inner product induced on the tangent plane to the surface by g. This inner product has the following coordinate matrix with respect to the basis $\frac{\partial q}{\partial x^1}$, $\frac{\partial q}{\partial x^2}$:

$$(g_{ij}) = \left(\sum_{k=1}^{3} \frac{\partial q^k}{\partial x^i} \frac{\partial q^k}{\partial x^j} \right)_{i,j=1}^{2}, \quad (g^{ij}) = (g_{ij})^{-1}, \quad \operatorname{vol}(g) = \sqrt{\det(g_{ij})} \, . \tag{6}$$

We denote by (g^{ij}) the inverse matrix of (g_{ij}) and by $\operatorname{vol}(g)$ the volume density of the surface q with respect to the metric g.

In this paper we will use the H^1-type inner metric on surfaces defined by the expression

$$\langle u, v\rangle_q := \sum_{k=1}^{3} \int_M u^k v^k + \alpha^2 g\left(\mathrm{grad}^g(u^k), \mathrm{grad}^g(v^k)\right) \mathrm{vol}(g)\, dx \tag{7}$$

$$= \sum_{k=1}^{3} \int_M u^k v^k + \alpha^2 g^{ij} \frac{\partial u^k}{\partial x^i} \frac{\partial v^k}{\partial x^j} \mathrm{vol}(g)\, dx \ . \tag{8}$$

One reason to use this generalization of the H^1-metric is, that this metric is invariant under reparametrizations of the surface and only depends on the image $q(M)$ as a subset of $\mathbb{R}^3$, in a similar way as the length of a curve in two dimensions only depends on the image of the curve, and not on a particular parametrization. This is necessary, if one wants to match unparametrized surfaces. For this task it is possible to use the same framework with this metric, only the matching term has to be chosen to beinvariant under reparametrizations.

The constant α, which appears in the metric, is a parameter, which has to be chosen for each problem. It represents the characteristic length scale, at which deformations take place. Another interpretation of α is the scale across which the momentum is smoothed, when passing from momenta to velocities. It can be compared with the kernel size in LDDMM.

Other choices for the metric are possible. One could use Sobolev type metrics of higher order as in [2] or multiply the components of the metric with a function depending on geometric quantities of the surface, like area, mean or Gaussian curvature as was done in [1].

2.5 The Matching Functional

There are different possible choices for the matching term. In this paper we will use the squared L^2-distance

$$d(q_0, q_1) = \int_M |q_0(x) - q_1(x)|^2\, dx. \tag{9}$$

Since we are dealing with parametrized surfaces this is a natural choice for the matching functional.

When matching unparametrized surfaces, natural choices of the matching functional would include currents, see [9], or one could use the reparametrization framework of [5].

3 Discretization

In this section we will describe how to discretize the variational problem (5) and compute the optimal path between two surfaces. Starting with an initial guess for the velocity u_0^1, we will use a gradient descent scheme

$$u_0^{i+1} = u_0^i - \varepsilon \nabla_{u_0} E(u_0^i)$$

to converge towards the initial velocity of the optimal geodesic. The discretization thus consists of two parts:

- compute the geodesic, given the initial velocity to evaluate $E(u_0^i)$
- compute the gradient $\nabla_{u_0} E(u_0^i)$ to update the initial velocity.

We show how to discretize the geodesic equation in Sec. 3.1 and how to compute the gradient in Sec. 3.2.

3.1 The Geodesic Equation

We discretize the time-evolution of the surface $q(t)$ using the explicit Euler method

$$q_{i+1} = q_i + \Delta t u_i \ , \tag{10}$$

where $q_i = q(i\Delta t)$ is the discretized version of the curve and $\Delta t = 1/N$ is the time step, if we divide the interval $[0, 1]$ into N parts. To compute u_i we note that a geodesic is a critical point of the energy $E(u_i) = \frac{\Delta t}{2} \sum_{i=0}^{N-1} \langle u_i, u_i \rangle_{q_i}$, i.e. $\nabla_{u_i} E(u_i) = 0$. Following [6] we introduce the Lagrangian multiplier p_i in the discrete variational principle

$$E(u_0, \dots, u_{N-1}) = \sum_{i=0}^{N-1} \frac{\Delta t}{2} \langle u_i, u_i \rangle_{q_i} + \langle p_i, q_{i+1} - q_i - \Delta t u_i \rangle_{L^2} \ . \tag{11}$$

and take variations. From variations in u_i we see that p_i is the momentum dual to the velocity u_i in the sense that $\langle u_i, \delta u_i \rangle_{q_i} = \langle p_i, \delta u_i \rangle_{L^2}$ and we obtain the evolution equation for u_i in the form

$$\langle u_{i+1}, \delta q_{i+1} \rangle_{q_{i+1}} = \langle u_i, \delta q_{i+1} \rangle_{q_i} + \Delta t \left\langle \frac{\delta \ell}{\delta q}(u_{i+1}, u_{i+1}; q_{i+1}), \delta q_{i+1} \right\rangle_{L^2} \ , \tag{12}$$

with δq_{i+1} an arbitrary variation. Here we use the notation

$$\ell(u, v; q) = \frac{1}{2} \langle u, v \rangle_q \ . \tag{13}$$

We denote by $\frac{\delta \ell}{\delta q}$ the variational derivative of $\ell(u, v; q)$ with respect to the variable q, defined via

$$\left\langle \frac{\delta \ell}{\delta q}(u, v; q), \delta q \right\rangle_{L^2} = \lim_{h \to 0} \frac{\ell(u, v; q + h\delta q) - \ell(u, v; q))}{h} \ . \tag{14}$$

Equation (12) is an implicit time step for u_{i+1}, since u_{i+1} appears on the right hand side in a quadratic term. To make computations easier and avoid having to solve a nonlinear equation, we changed the right hand side to an explicit Euler time step

$$\langle u_{i+1}, \delta q \rangle_{q_{i+1}} = \langle u_i, \delta q \rangle_{q_i} + \Delta t \left\langle \frac{\delta \ell}{\delta q}(u_i, u_i; q_i), \delta q \right\rangle_{L^2} \ . \tag{15}$$

3.2 Computing the Gradient

Given (10), (15) for the evolution of a geodesic we again use method of adjoint equations from [6] to compute the gradient of the energy with respect to the initial velocity. The resulting equations for the variables $\widehat{u}_i, \widehat{v}_i$ have to be integrated backwards in time

$$\begin{aligned}
\langle \widehat{u}_i, \delta u_i \rangle_{q_i} &= \langle \widehat{u}_{i+1}, \delta u_i \rangle_{q_i} + \Delta t \langle \widehat{v}_{i+1}, \delta u_i \rangle_{q_{i+1}} + 2\Delta t \langle \frac{\delta \ell}{\delta q}(u_i, \delta u_i; q_i), \widehat{u}_{i+1} \rangle \\
\langle \widehat{v}_i, \delta q_i \rangle_{q_i} &= \langle \widehat{v}_{i+1}, \delta q_i \rangle_{q_{i+1}} + 2 \left\langle \frac{\delta \ell}{\delta q}(\widehat{u}_{i+1} - \widehat{u}_i, u_i; q_i), \delta q_i \right\rangle \\
&\quad + \Delta t \left\langle \frac{\delta^2 \ell}{\delta q^2}(u_i, u_i; q_i), (\widehat{u}_{i+1}, \delta q_i) \right\rangle
\end{aligned} \tag{16}$$

with the initial conditions

$$\widehat{u}_N = 0 \qquad \langle \widehat{v}_N, \delta q_N \rangle_{q_N} = -\frac{1}{\sigma^2} \langle q_N - q_{\text{targ}}, \delta q_N \rangle \tag{17}$$

at time $t = 1$. The gradient is then given by

$$\nabla_{u_0} E(u_0) = u_0 - \widehat{u}_0 \ . \tag{18}$$

4 Numerical Experiments

We implemented the geodesic and adjoint equations (10), (15) and (16) in Python using the finite element library FEniCS [14]. All model manifolds ($[0,1] \times [0,1]$, $S^1 \times [0,1]$, $S^1 \times S^1$) were modelled on the rectangle $[0,1] \times [0,1]$ with periodic boundary conditions prescribed where neccessary. The domain was subdivided into a regular triangular mesh, on which Lagrangian finite elements of order 1 were defined.

In the first example we apply our method to compute the geodesic path between two shapes, which includes both large and small deformations. The template shape is a straight cylinder of height 1 and radius 0.25, which is discretized using a regular triangular mesh of $2 \times 30 \times 30$ elements. The target shape is a cylinder, which is bent by 90° and has 5 small ripples added to it along the vertical axis. Compared to the bending the ripples constitute a small and local deformation of the shape. The target shape is discretized in the same way as the template. We use $\alpha = 0.6$ as the length scale parameter and 10 time steps for the time integration. The gradient descent takes 80 steps to converge to an L^2-error of 0.008. We can see in Fig. 2 that both the large and the small deformations are captured by the geodesic.

In the second example we want to illustrate the curved nature of shape space. To do so we pick three asymmetric tori, lying in different positions in space. Each two tori differ by a composition of two rigid rotations. We compute the geodesics between each pair of tori to measure the angles and side lengths of the triangle with the tori as vertices and the geodesics as edges. By comparing the sum of the

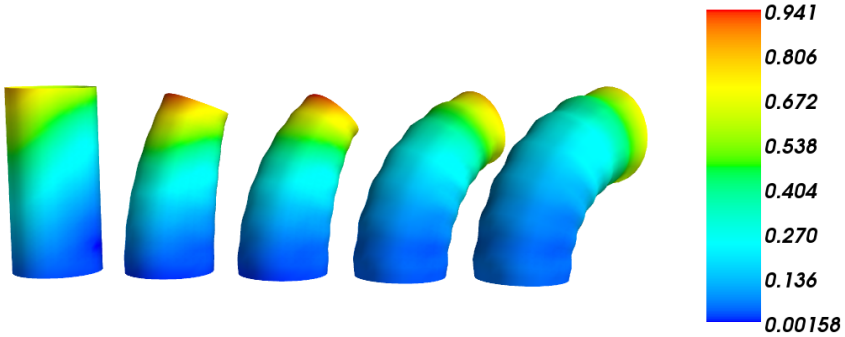


Fig. 2. Samples are shown from a geodesic in the space of surfaces between a straight cylinder and a bent cylinder with ripples, at time points $t = 0, 0.3, 0.5, 0.8, 1$. The color encodes the Euclidean length of the deformation vector field at each point of the surface.

angles with π one can estimate, whether the curvature of shape space along the plane containing the triangle is positive or negative. We measured $\alpha = 33.766°$, $\beta = 34.802°$ and $\gamma = 34.675°$. The sum $\alpha + \beta + \gamma = 103.243°$ is smaller than $180°$, which indicates that the space is negatively curved in this area (c.f. [16, Sec. 5.4]). In negatively curved spaces geodesics tend to be attracted towards a common point. In this example the geodesics are attracted towards the surface, which is degenerated to a point. We can see in Fig. 3 that the midpoints of the geodesics between the vertices are slightly shrunk. This is another indication for the negatively curved nature of the space.

In the third example we show, that our framework is indeed capable to do nonlinear statistics on shape space. We generate five sample shapes and compute the mean shape between them. The five shapes are cylindrical vases with an open top and bottom, discretized again using a triangular mesh of $2 \times 30 \times 30$ elements. As the initial guess for the mean we use a straight cylinder. First we register this initial shape to the five target shapes, compute the average of the initial velocities and then shoot with this average velocity to obtain a next guess for the mean shape. We iterate this procedure until the average velocity is close to zero. This method of computing the Karcher mean was proposed in [7]. After four iterations we obtained an average velocity with norm 0.006. As can be seen in Fig. 4 the average shape indeed combines the characteristics of the five shapes.

5 Conclusions

In this paper we propose a metric to match regular surfaces in a Riemannian setting. Although this metric has been studied from a mathematical point of view in several papers, including [2, 17], so far it hasn't been applied to problems in computational anatomy. The aim of this work is to argue, that this is a promising approach, which is worthwhile to be studied further.

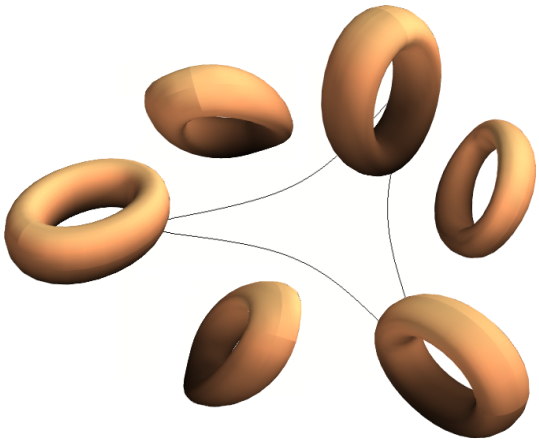


Fig. 3. This figure shows a geodesic triangle in the space of surfaces with asymmetric tori as vertices. The tori along the edges are the middle points of the geodesics connecting the vertices. One can see that the shapes tend to shrink along the geodesics before expanding again towards the ends. This effect implies negative curvature in this region of shape space.

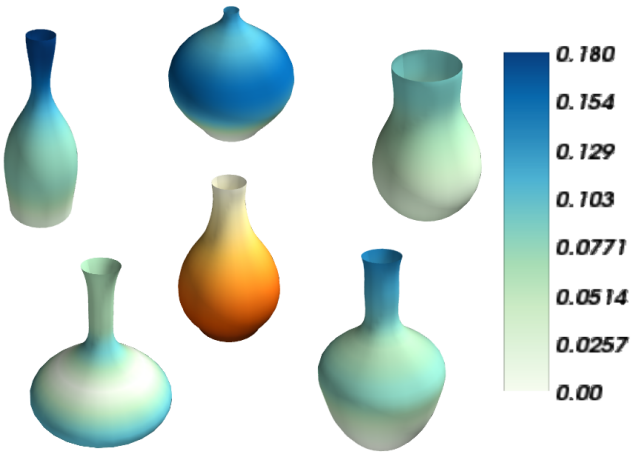


Fig. 4. In this figure we show the Karcher mean of five vase-shaped objects. The mean shape, which is displayed in the center of the figure is computed using an iterated shooting method. The colored regions on the averaged shapes encode the Euclidean length of the initial velocity of the geodesic, which connects each shape to the mean. The color of the mean was chosen for artistic purposes only.

In contrast to LDDMM, where surfaces are deformed via a deformation of the ambient space, in this approach the deformation is prescribed directly on the surface, while the ambient space stays constant. Because of this we call this approach matching with *inner metrics* as opposed to LDDMM, which can be described as matching with *outer metrics*. Other inner metrics, which have been proposed in the literature include [18, 12, 23, 8, 1] We show how to discretize the geodesic equation and how to compute the gradient of the matching functional with respect to the initial velocity. In the last part of the paper we present numerical results on synthetic data of different topologies, which demonstrate the versatility and applicability of our method.

At the present we applied this metric to match parametrized surfaces, which is an unwelcome restriction in applications. This is not a restriction of the framework itself, but only of the matching functional. By choosing a matching functional, which is independent of the parametrization of the surface, one can apply the same framework to unparametrized surfaces. In future work we plan to implement this capability and test the method on real anatomical data.

Acknowledgments

We are enormously grateful to our friends and colleagues Colin Cotter, Phillip Harms, Darryl Holm, Peter Michor and David Mumford for their thoughtful comments and encouragement in the course of this work.

We also acknowledge partial support by the Royal Society of London Wolfson Award, the European Research Council Advanced Grant, the Imperial College London SIF Programme and the Austrian Science Fund (FWF).

References

1. Bauer, M., Harms, P., Michor, P.W.: Almost local metrics on shape space of hypersurfaces in n-space. (Submitted. Preprint available at arXiv:1001.0717) (2010)
2. Bauer, M., Harms, P., Michor, P.W.: Sobolev metrics on shape space of surfaces in n-space. (To appear in Journal of Geometric Mechanics.) (2011)
3. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vision 61(2), 139–157 (2005)
4. Bruveris, M., Gay-Balmaz, F., Holm, D., Ratiu, T.: The momentum map representation of images. Journal of Nonlinear Science 21, 115–150 (2011)
5. Cotter, C.J., Holm, D.D.: Geodesic boundary value problems with symmetry. J. Geom. Mech. 2(1), 51–68 (2010)
6. Dimet, F.X.L., Ngodock, H.E., Navon, I.M.: Sensitivity analysis in variational data assimilation. J. Meteorol. Soc. Japan pp. 145–155 (1997)
7. Fletcher, P., Venkatasubramanian, S., Joshi, S.: Robust statistics on riemannian manifolds via the geometric median. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1 –8 (june 2008)
8. Fuchs, M., Jüttler, B., Scherzer, O., Yang, H.: Shape metrics based on elastic deformations. J. Math. Imaging Vis. 35, 86–102 (September 2009)

9. Glaunès, J., Qiu, A., Miller, M.I., Younes, L.: Large deformation diffeomorphic metric curve mapping. Int. J. Comput. Vision 80(3) (2008)
10. Grenander, U.: General Pattern Theory. Oxford Science Publications (1993)
11. Kilian, M., Mitra, N.J., Pottmann, H.: Geometric modeling in shape space. ACM Trans. Graph. 26 (July 2007)
12. Kurtek, S., Klassen, E., Ding, Z., Srivastava, A.: A novel riemannian framework for shape analysis of 3d objects. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 0, 1625–1632 (2010)
13. Liu, X., Shi, Y., Dinov, I., Mio, W.: A computational model of multidimensional shape. Int. J. Comput. Vision 89, 69–83 (August 2010)
14. Logg, A., Wells, G.N.: Dolfin: Automated finite element computing. ACM Trans. Math. Softw. 37, 20:1–20:28 (April 2010),
15. Bertalmio, M., Cheng, L-T., Osher, S., Sapiro, G.: Variational problems and partial differential equations on implicit surfaces. Journal of Computational Physics 174(2), 759–780 (2001)
16. Michor, P.W., Mumford, D.: Riemannian geometries on spaces of plane curves. J. Eur. Math. Soc. (JEMS) 8 (2006), 1-48 (2006),
17. Michor, P.W., Mumford, D.: An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. Appl. Comput. Harmon. Anal. 23(1), 74–113 (2007)
18. Mio, W., Srivastava, A., Joshi, S.: On shape of plane elastic curves. International Journal of Computer Vision 73, 307–324 (2007)
19. Mmoli, F., Sapiro, G., Osher, S.: Solving variational problems and partial differential equations mapping into general target manifolds. Journal of Computational Physics 195(1), 263 – 292 (2004)
20. Peter, A., Rangarajan, A.: Maximum likelihood wavelet density estimation with applications to image and shape matching. Image Processing, IEEE Transactions on 17(4), 458 –468 (april 2008)
21. Risser, L., Vialard, F.X., Murgasova, M., Holm, D., Rueckert, D.: Large deformation diffeomorphic registration using fine and coarse strategies. In: Fischer, B., Dawant, B., Lorenz, C. (eds.) Biomedical Image Registration, Lecture Notes in Computer Science, vol. 6204, pp. 186–197. Springer Berlin / Heidelberg (2010)
22. Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. (Submitted.) (2011)
23. Wirth, B., Bar, L., Rumpf, M., Sapiro, G.: A continuum mechanical approach to geodesics in shape space. IJCV 93(3), 293–318 (2011)
24. Younes, L., Michor, P.W., Shah, J., Mumford, D.: A metric on shape space with explicit geodesics. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 19(1), 25–57 (2008)
25. Younes, L., Arrate, F., Miller, M.I.: Evolutions equations in computational anatomy. NeuroImage 45(1, Supplement 1), S40 – S50 (2009), mathematics in Brain Imaging

Author Index